

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: May 12, 2003, 10:00:31 ; Search time 40 Seconds  
(without alignments)  
1748.914 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 525  
Sequence: 1 MRANPKYQAMPSETISIMKT.....KLWKRIIAKLISLPIRESLL 525

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A\_Geneseq\_101002.\*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	508	96.8	508	21	Neisseria meningit
2	237	45.1	508	21	Neisseria meningit
3	215	41.0	507	21	Neisseria meningit
4	8	1.5	71	23	Neisseria gonorrhoe
5	8	1.5	100	22	Human ORF3243 prot
6	8	1.5	498	22	Protonibacterium
7	8	1.5	579	22	C glutamicum prote
8	8	1.5	584	22	Protein encoded by
9	8	1.5	683	22	Corynebacterium gl
10	8	1.5	692	22	Novel human diagno

11	8	1.5	796	22	ABG01448	Novel human diagno
12	8	1.5	796	22	ABG11797	Novel human diagno
13	8	1.5	882	22	ABG05821	Novel human diagno
14	8	1.5	1275	12	AA13516	P. denitrificans CO
15	7	1.3	18	15	AA19463	HLA-DR alpha-chain
16	7	1.3	19	15	AA19462	HLA-DR alpha-chain
17	7	1.3	21	13	AA19462	HLA-DR alpha-chain
18	7	1.3	21	13	AA19462	HLA-DR alpha-chain
19	7	1.3	21	14	AA19462	HLA-DR alpha-chain
20	7	1.3	21	14	AA19462	HLA-DR alpha-chain
21	7	1.3	21	15	AA19462	HLA-DR alpha-chain
22	7	1.3	21	15	AA19462	HLA-DR alpha-chain
23	7	1.3	21	15	AA19462	HLA-DR alpha-chain
24	7	1.3	21	15	AA19462	HLA-DR alpha-chain
25	7	1.3	21	15	AA19462	HLA-DR alpha-chain
26	7	1.3	21	15	AA19462	HLA-DR alpha-chain
27	7	1.3	21	15	AA19462	HLA-DR alpha-chain
28	7	1.3	21	15	AA19462	HLA-DR alpha-chain
29	7	1.3	21	15	AA19462	HLA-DR alpha-chain
30	7	1.3	21	15	AA19462	HLA-DR alpha-chain
31	7	1.3	21	15	AA19462	HLA-DR alpha-chain
32	7	1.3	21	15	AA19462	HLA-DR alpha-chain
33	7	1.3	21	15	AA19462	HLA-DR alpha-chain
34	7	1.3	21	15	AA19462	HLA-DR alpha-chain
35	7	1.3	21	15	AA19462	HLA-DR alpha-chain
36	7	1.3	21	15	AA19462	HLA-DR alpha-chain
37	7	1.3	21	15	AA19462	HLA-DR alpha-chain
38	7	1.3	21	15	AA19462	HLA-DR alpha-chain
39	7	1.3	21	15	AA19462	HLA-DR alpha-chain
40	7	1.3	21	15	AA19462	HLA-DR alpha-chain
41	7	1.3	21	15	AA19462	HLA-DR alpha-chain
42	7	1.3	21	15	AA19462	HLA-DR alpha-chain
43	7	1.3	21	15	AA19462	HLA-DR alpha-chain
44	7	1.3	21	15	AA19462	HLA-DR alpha-chain
45	7	1.3	21	15	AA19462	HLA-DR alpha-chain
46	7	1.3	21	15	AA19462	HLA-DR alpha-chain
47	7	1.3	21	15	AA19462	HLA-DR alpha-chain
48	7	1.3	21	15	AA19462	HLA-DR alpha-chain
49	7	1.3	21	15	AA19462	HLA-DR alpha-chain
50	7	1.3	21	15	AA19462	HLA-DR alpha-chain
51	7	1.3	21	15	AA19462	HLA-DR alpha-chain
52	7	1.3	21	15	AA19462	HLA-DR alpha-chain
53	7	1.3	21	15	AA19462	HLA-DR alpha-chain
54	7	1.3	21	15	AA19462	HLA-DR alpha-chain
55	7	1.3	21	15	AA19462	HLA-DR alpha-chain
56	7	1.3	21	15	AA19462	HLA-DR alpha-chain
57	7	1.3	21	15	AA19462	HLA-DR alpha-chain
58	7	1.3	21	15	AA19462	HLA-DR alpha-chain
59	7	1.3	21	15	AA19462	HLA-DR alpha-chain
60	7	1.3	21	15	AA19462	HLA-DR alpha-chain
61	7	1.3	21	15	AA19462	HLA-DR alpha-chain
62	7	1.3	21	15	AA19462	HLA-DR alpha-chain
63	7	1.3	21	15	AA19462	HLA-DR alpha-chain
64	7	1.3	21	15	AA19462	HLA-DR alpha-chain
65	7	1.3	21	15	AA19462	HLA-DR alpha-chain
66	7	1.3	21	15	AA19462	HLA-DR alpha-chain
67	7	1.3	21	15	AA19462	HLA-DR alpha-chain
68	7	1.3	21	15	AA19462	HLA-DR alpha-chain
69	7	1.3	21	15	AA19462	HLA-DR alpha-chain
70	7	1.3	21	15	AA19462	HLA-DR alpha-chain
71	7	1.3	21	15	AA19462	HLA-DR alpha-chain
72	7	1.3	21	15	AA19462	HLA-DR alpha-chain
73	7	1.3	21	15	AA19462	HLA-DR alpha-chain
74	7	1.3	21	15	AA19462	HLA-DR alpha-chain
75	7	1.3	21	15	AA19462	HLA-DR alpha-chain
76	7	1.3	21	15	AA19462	HLA-DR alpha-chain
77	7	1.3	21	15	AA19462	HLA-DR alpha-chain
78	7	1.3	21	15	AA19462	HLA-DR alpha-chain
79	7	1.3	21	15	AA19462	HLA-DR alpha-chain
80	7	1.3	21	15	AA19462	HLA-DR alpha-chain
81	7	1.3	21	15	AA19462	HLA-DR alpha-chain
82	7	1.3	21	15	AA19462	HLA-DR alpha-chain
83	7	1.3	21	15	AA19462	HLA-DR alpha-chain

Novel human diagno  
Novel human diagno  
Novel human diagno  
P. denitrificans CO  
HLA-DR alpha-chain  
HLA-DR alpha-chain  
Thrl8, Leu19 Endoc  
Endothelin recepto  
Endothelin metal c  
Endothelin metal c  
Endothelin-2. Syn  
Endothelin peptide  
Endothelin peptide  
Endothelin peptide  
Endothelin peptide  
Endothelin peptide  
Endothelin peptide  
Peptide ligand #9  
Peptide ligand #2  
[Thrl8, Leu19]-ET-  
Endothelin sequenc  
Endothelin recepto  
Targeting conjuga  
Targeting conjuga  
Targeting conjuga  
Conjugating peptid  
Endothelin recepto  
Endothelin recepto  
Endothelin recepto  
Endothelins and re  
HLA-DR alpha-chain  
Maize ZmKCS1 pepti  
Peptide #8100 enco  
Protein #6883 enco  
Human bone marrow  
Human bone marrow  
Peptide #8391 enco  
Human peptide enco  
Endothelin-2 gene  
Synthetic hET2. S  
Human big endothel  
Endothelins and re  
Endothelins and re  
Human secreted pro  
Human secreted pro  
Human Immune/Phema  
Human colon cancer  
Human secreted pro  
Drosophila melanog  
Streptococcus poly  
Human leukocyte su  
Human Thiolase-11k  
Human ORF3 protein  
Zea mays protein f  
Human protein sequ  
Human ORF3 protein  
Protonibacterium  
Expressed antigen  
Novel human diagno  
H. pylori GPO 111  
Antigen 1 from clu  
Human secreted pro  
C glutamicum prote  
Sinorhizobium mell  
Human polypeptide  
Novel human polype  
Human endothelin-2  
Precursor ET-2 seq  
C. elegans LIN-8 r

84	7	1.3	178	23	AAE15748	Human endothelin 2
85	7	1.3	186	22	ABG01371	Novel human diagno
86	7	1.3	191	23	AAU09849	Novigen presenting
87	7	1.3	205	22	ABB96331	Human testicular a
88	7	1.3	205	22	AAAB5800	Human reproductive
89	7	1.3	221	22	ABG20012	Novel human diagno
90	7	1.3	229	16	AAK74037	MHC polypeptide HL
91	7	1.3	235	20	AAK37024	Chlamydia trachoma
92	7	1.3	237	22	ABB61036	Drosophila melanog
93	7	1.3	237	22	AAAM42088	Human polypeptide
94	7	1.3	238	22	AAU65742	Propionibacterium
95	7	1.3	245	19	AAAM6943	DR-alpha extracell
96	7	1.3	245	20	AAK31652	HLA-DR2 alpha-Fos
97	7	1.3	246	21	AAK22115	Maize glutathione-
98	7	1.3	248	18	AAW37341	DR alpha-DNA chime
99	7	1.3	252	23	ABP40674	Staphylococcus epi
100	7	1.3	253	21	AAV68277	Class II alpha cha
101	7	1.3	253	21	AAV52931	Class II alpha cha
102	7	1.3	253	22	AAV58692	Class II alpha cha
103	7	1.3	254	22	ABB50276	HLA-DR alpha chain
104	7	1.3	254	23	AAK79361	Human HLA DR alpha
105	7	1.3	259	21	AAV75333	Neisseria meningit
106	7	1.3	261	21	AAAG10024	Arabidopsis thalia
107	7	1.3	261	21	AAAG27624	Arabidopsis thalia
108	7	1.3	272	23	AAU73403	Ehrlichia chaffeen
109	7	1.3	281	21	AAV58160	Lung cancer associ
110	7	1.3	281	22	AAK82399	S. epidermidis ope
111	7	1.3	285	21	AAAB03446	Canidia albicans e
112	7	1.3	298	22	AAU36841	Staphylococcus aur
113	7	1.3	299	21	AAAG50128	Arabidopsis thalia
114	7	1.3	302	21	AAV56657	Human prostate can
115	7	1.3	303	22	AAK82023	S. epidermidis ope
116	7	1.3	306	21	AAK24844	Arabidopsis thalia
117	7	1.3	306	21	AAK45182	Arabidopsis thalia
118	7	1.3	307	21	AAK45181	Arabidopsis thalia
119	7	1.3	307	21	AAK45181	Arabidopsis thalia
120	7	1.3	310	21	AAK24842	Arabidopsis thalia
121	7	1.3	310	21	AAK45180	Arabidopsis thalia
122	7	1.3	318	22	AAW94021	Human stomach can
123	7	1.3	318	22	AAK95783	Human protein sequ
124	7	1.3	318	23	AAW49036	Human phosphatidy
125	7	1.3	320	22	AAU34154	Staphylococcus aur
126	7	1.3	324	21	AAK14143	Borrelia pertuss
127	7	1.3	327	21	AAK19279	Amino acid sequen
128	7	1.3	327	22	AAK60877	Mevlonate pathway
129	7	1.3	353	22	ABG00630	Novel human diagno
130	7	1.3	359	22	AAK82824	S. epidermidis ope
131	7	1.3	375	21	AAAG10023	Arabidopsis thalia
132	7	1.3	375	21	AAK27623	Arabidopsis thalia
133	7	1.3	380	21	AAK31232	Arabidopsis thalia
134	7	1.3	380	21	AAAG50119	Arabidopsis thalia
135	7	1.3	380	21	AAAG50127	Arabidopsis thalia
136	7	1.3	381	22	ABG15205	Novel human diagno
137	7	1.3	384	21	AAK21817	Arabidopsis thalia
138	7	1.3	385	21	ABG03415	Novel human diagno
139	7	1.3	387	22	ABG26063	Novel human diagno
140	7	1.3	404	21	AAK31231	Arabidopsis thalia
141	7	1.3	404	21	AAAG50118	Arabidopsis thalia
142	7	1.3	404	21	AAAG50126	Arabidopsis thalia
143	7	1.3	413	22	AAAG90724	C glutaminc prote
144	7	1.3	413	22	AAAB59687	C. elegans KAS pro
145	7	1.3	415	22	ABB66250	Drosophila melanog
146	7	1.3	418	22	AAE13040	R. sanguineus glut
147	7	1.3	427	22	ABG29887	Novel human diagno
148	7	1.3	428	19	AAW98378	H. pylori GHP 137
149	7	1.3	430	23	ABP27391	Streptococcus poly
150	7	1.3	433	21	AAAG50117	Arabidopsis thalia
151	7	1.3	441	22	AAK79050	Corynebacterium q1
152	7	1.3	449	22	AAK68879	Human RECAP polype
153	7	1.3	453	21	AAK94340	Human cell surface
154	7	1.3	458	23	ABP27354	Streptococcus poly
155	7	1.3	462	22	ABB66252	Drosophila melanog
156	7	1.3	467	16	AAK83968	GABA-A receptor ga
157	7	1.3	474	22	ABG06097	Novel human diagno
158	7	1.3	474	23	ABK07495	Human lipid metabo
159	7	1.3	477	21	AAK31230	Arabidopsis thalia
160	7	1.3	477	23	ABK32097	Hericidially activ
161	7	1.3	479	20	AAK31654	HLA-DR2 alpha-Fos
162	7	1.3	486	23	AAE22711	Human transporter
163	7	1.3	487	22	AAK39684	Human protein sequ
164	7	1.3	489	23	AAE12074	Dendritic cell (DC
165	7	1.3	489	23	AAU86143	Human PRO5775 poly
166	7	1.3	489	23	ABG34050	Human pro peptide
167	7	1.3	494	21	AAK45012	Maize zmks01 prote
168	7	1.3	496	19	AAK37059	HiV-1 breakthrough
169	7	1.3	499	22	AAU46828	Propionibacterium
170	7	1.3	502	18	AAW55452	H. pylori ORF 02ae
171	7	1.3	502	18	AAW55550	Listeria monocytog
172	7	1.3	504	23	ABK49813	Sequence encoded b
173	7	1.3	505	9	AAK81265	Human secreted pro
174	7	1.3	506	21	AAU94986	Human phospholips
175	7	1.3	506	23	AAU99377	Corn raffinose syn
176	7	1.3	508	21	AAK70974	Drosophila melanog
177	7	1.3	508	22	ABK58762	Arabidopsis thalia
178	7	1.3	514	21	AAK21397	Arabidopsis thalia
179	7	1.3	514	22	ABK60741	Drosophila melanog
180	7	1.3	514	22	ABK60742	Drosophila melanog
181	7	1.3	515	22	AAK94919	Human protein sequ
182	7	1.3	516	23	AAK22913	Human transporter
183	7	1.3	521	20	AAW30638	Partial human 7-tr
184	7	1.3	525	20	AAK25881	Human secreted pro
185	7	1.3	525	21	AAK21396	Arabidopsis thalia
186	7	1.3	542	23	ABK53329	Lactococcus lactis
187	7	1.3	543	22	AAK83170	Corynebacterium th
188	7	1.3	546	21	AAK21395	Arabidopsis thalia
189	7	1.3	573	22	AAK97498	Murine DNA polymer
190	7	1.3	575	22	AAW40203	Human polypeptide
191	7	1.3	575	22	AAK97499	Human DNA polymera
192	7	1.3	581	22	AAK93657	Human polypeptide,
193	7	1.3	581	23	ABK77177	Human oxidoreducta
194	7	1.3	585	21	AAK53078	Arabidopsis thalia
195	7	1.3	585	21	AAK54311	Arabidopsis thalia
196	7	1.3	601	17	AAK88405	Human DNA-topoisom
197	7	1.3	601	19	AAW41362	Human G-protein co
198	7	1.3	611	23	AAK18647	HLA-DR2 alpha-Fos-
199	7	1.3	612	20	AAK16655	Human 7-transmembr
200	7	1.3	652	20	AAK30637	Human EGFR polype
201	7	1.3	652	21	AAV59300	Human expressed po
202	7	1.3	661	22	ABK03062	Human CDNA seq ID
203	7	1.3	661	22	ABK10371	Human novel uterin
204	7	1.3	661	22	AAU18138	Human endocrine po
205	7	1.3	661	22	AAU18399	Human novel secret
206	7	1.3	661	22	AAU16983	Novel human diagno
207	7	1.3	682	22	ABG03410	Novel human diagno
208	7	1.3	682	22	ABG08044	C glutaminc prote
209	7	1.3	695	22	AAK91773	Adhesin. Escheric
210	7	1.3	696	17	AAK75366	Arabidopsis thalia
211	7	1.3	718	21	AAK43549	Novel human diagno
212	7	1.3	723	22	ABG01373	Arabidopsis thalia
213	7	1.3	736	21	AAK43548	Streptococcus poly
214	7	1.3	753	23	ABP29819	Drosophila melanog
215	7	1.3	761	23	ABP27624	Hericidially activ
216	7	1.3	765	22	ABK62680	Hericidially activ
217	7	1.3	773	21	AAK43547	Hericidially activ
218	7	1.3	773	23	ABK92720	Hericidially activ
219	7	1.3	776	23	ABK91316	Hericidially activ
220	7	1.3	780	22	ABG02105	Novel human diagno
221	7	1.3	780	22	ABG09509	Novel human diagno
222	7	1.3	800	23	AAE13609	Human mature EGF-1
223	7	1.3	813	22	ABK58606	Drosophila melanog
224	7	1.3	813	22	ABK58606	Drosophila melanog
225	7	1.3	815	22	ABK60075	Drosophila melanog
226	7	1.3	823	23	AAK13608	Human EGF-like mol
227	7	1.3	1018	21	AAK25534	Eucalyptus grandis
228	7	1.3	1021	18	AAK23281	Allorreaction assoc
229	7	1.3	1074	22	AAU40997	Propionibacterium



230	7	1.3	1194	22	ABR62608	Drosophila melanog	303	6	1.1	25	22	AA92159	Signal transductio
231	7	1.3	1207	21	AA94746	Amino acid sequenc	304	6	1.1	25	22	AA964685	Human secreted pro
232	7	1.3	1207	21	AA94747	Amino acid sequenc	305	6	1.1	25	19	AA937984	Amino acid sequenc
233	7	1.3	1207	21	AA94748	Amino acid sequenc	306	6	1.1	26	20	AA94318	Human pUr-alpha cl
234	7	1.3	1207	21	AA94749	Amino acid sequenc	307	6	1.1	26	20	AA94318	Modified surfactan
235	7	1.3	1208	21	AA94742	Amino acid sequenc	308	6	1.1	27	22	AA94692	Human surfactant p
236	7	1.3	1208	21	AA94743	Amino acid sequenc	309	6	1.1	27	22	AA946109	Human surfactant p
237	7	1.3	1208	21	AA94744	Amino acid sequenc	310	6	1.1	28	14	AA930660	Peptide contg. alt
238	7	1.3	1208	21	AA94745	Amino acid sequenc	311	6	1.1	28	20	AA981197	Surfactant peptide
239	7	1.3	1208	21	AA932220	Human TIM protein	312	6	1.1	28	22	AB940888	Peptide #8394 enco
240	7	1.3	1317	22	AA936222	Pseudomonas aerugi	313	6	1.1	28	22	AB941240	Peptide #8746 enco
241	7	1.3	1498	23	AB948222	Yersinia monocytog	314	6	1.1	28	22	AB925027	Protein #7026 enco
242	7	1.3	1603	22	AA940302	Human polypeptide	315	6	1.1	28	22	AB925239	Protein #7238 enco
243	7	1.3	2071	21	AA94686	Amino acid sequenc	316	6	1.1	28	22	AA961748	Human brain expres
244	7	1.3	2071	22	AA911052	Amino acid sequenc	317	6	1.1	28	22	AA962106	Human brain expres
245	7	1.3	2771	22	AB938880	Asbysa gossypii GT	318	6	1.1	28	22	AA974544	Human bone marrow
246	7	1.3	3070	22	AB938195	Drosophila melanog	319	6	1.1	28	22	AA974909	Human bone marrow
247	7	1.3	6239	21	AA937750	Streptomyces averm	320	6	1.1	28	22	AA940338	Peptide #6772 enco
248	7	1.3	6239	22	AA932251	Streptomyces averm	321	6	1.1	28	22	AA94656	Peptide #8693 enco
249	6	1.1	9	21	AA932345	HIV B2702 binding	322	6	1.1	28	23	AB944398	Peptide #9061 enco
250	6	1.1	9	22	AA968933	Human TADG-12 immu	323	6	1.1	28	23	AB944668	Human peptide enco
251	6	1.1	9	22	AA969027	Human TADG-12 immu	324	6	1.1	28	23	AA989388	Human peptide enco
252	6	1.1	10	22	AA968758	Saccharomyces cere	325	6	1.1	29	21	AA956188	Insulin/Insulin-11
253	6	1.1	10	22	AA968759	Saccharomyces cere	326	6	1.1	31	20	AA963078	Arabidopsis thalia
254	6	1.1	11	21	AA938686	HIV-1 peptide epit	327	6	1.1	31	21	AA920889	Peptide from the 3
255	6	1.1	11	21	AA938687	HIV-1 peptide epit	328	6	1.1	31	21	AA920889	Human Lhx3b immuno
256	6	1.1	11	21	AA938688	HIV-1 peptide epit	329	6	1.1	31	21	AA956463	Porcine Lhx3b prot
257	6	1.1	11	21	AA938689	HIV-1 peptide epit	330	6	1.1	31	22	AB930091	Peptide #2742 enco
258	6	1.1	11	21	AA938690	HIV-1 peptide epit	331	6	1.1	31	22	AB930091	Peptide #2768 enco
259	6	1.1	11	21	AA938691	HIV-1 peptide epit	332	6	1.1	31	22	AB930091	Protein #2703 enco
260	6	1.1	11	21	AA938692	HIV-1 peptide epit	333	6	1.1	31	22	AA968093	Human bone marrow
261	6	1.1	13	20	AA948700	Membrane dipeptid	334	6	1.1	31	22	AA968463	Peptide #2708 enco
262	6	1.1	14	15	AA949854	Sequence of antige	335	6	1.1	31	22	AA968463	Human bone marrow
263	6	1.1	14	15	AA970023	Tick carboxypeptid	336	6	1.1	31	22	AA968760	Peptide #2797 enco
264	6	1.1	14	22	AA967739	Human SNP related	337	6	1.1	31	22	AA968760	Peptide #2689 enco
265	6	1.1	14	22	AA967739	Human peptide #14	338	6	1.1	31	22	AB938044	Human peptide enco
266	6	1.1	14	22	AA968736	Human peptide #131	339	6	1.1	31	22	AA985033	Peptide r111 from
267	6	1.1	14	22	AA968736	PC1, recognisid by	340	6	1.1	33	16	AA987200	Peptide r111 from
268	6	1.1	14	22	AA968736	Protein C inhibito	341	6	1.1	36	19	AA937387	eEF2 derived prese
269	6	1.1	16	21	AA976757	Novel recombinant	342	6	1.1	36	22	AA989235	Hepatitis C virus
270	6	1.1	16	21	AA934578	Human secreted pro	343	6	1.1	37	16	AA985021	Human immune/thema
271	6	1.1	16	22	AA904751	Human 11CBY peptid	344	6	1.1	37	16	AA985788	Peptide r111 from
272	6	1.1	18	22	AA988922	HIV gp120 protein	345	6	1.1	37	21	AA951401	Human secreted pep
273	6	1.1	18	22	AA988923	HIV gp120 protein	346	6	1.1	37	21	AA94953	Human 5' EST relat
274	6	1.1	18	22	AA988924	HIV gp120 protein	347	6	1.1	38	13	AA927638	Endothelin antago
275	6	1.1	19	14	AA937943	HCV NS-4 type 2 re	348	6	1.1	38	13	AA927638	Endothelin antago
276	6	1.1	19	14	AA930658	Peptide contg. alt	349	6	1.1	38	22	AA955873	Human immune/thema
277	6	1.1	19	20	AA988195	Surfactant peptide	350	6	1.1	39	13	AA927671	Endothelin antago
278	6	1.1	19	22	AA949690	Modified surfactan	351	6	1.1	39	13	AA927671	Endothelin antago
279	6	1.1	19	22	AA949690	Human surfactant p	352	6	1.1	40	16	AA94618	Hepatitis C virus
280	6	1.1	20	16	AA974614	Hepatitis C virus	353	6	1.1	40	21	AA940168	Human secreted pro
281	6	1.1	20	18	AA942903	Immunogenic Hepati	354	6	1.1	42	21	AA953062	Arabidopsis thalia
282	6	1.1	20	18	AA942903	Immunogenic Hepati	355	6	1.1	42	22	AB906138	Novel human diagno
283	6	1.1	20	19	AA94670	Mutant muscarinic	356	6	1.1	44	20	AA988562	Secreted protein e
284	6	1.1	20	20	AA930804	Amino acid sequenc	357	6	1.1	44	22	AB950329	Human secreted pro
285	6	1.1	20	20	AA930804	Mutant m5 muscarin	358	6	1.1	44	22	AA948114	Human RANGO 204 po
286	6	1.1	20	22	AA983107	Peptide from the 3	359	6	1.1	45	18	AA928327	Staphylococcus aur
287	6	1.1	20	22	AA983107	Amino acid sequenc	360	6	1.1	45	20	AA911484	Human 5' EST secre
288	6	1.1	21	12	AA912590	Platelet adhesio	361	6	1.1	45	21	AA956242	Human secreted pro
289	6	1.1	21	13	AA927596	Endothelin antago	362	6	1.1	45	21	AA976348	Fragment of human
290	6	1.1	21	13	AA927597	Endothelin antago	363	6	1.1	45	21	AA976348	Human prostate can
291	6	1.1	21	14	AA930659	Peptide contg. alt	364	6	1.1	45	22	AA949648	Human reproductive
292	6	1.1	21	14	AA930659	Surfactant peptide	365	6	1.1	45	22	AA966150	Human reproductive
293	6	1.1	21	22	AA969419	Synthetic HAV VP3	366	6	1.1	46	18	AA927955	Staphylococcus aur
294	6	1.1	21	22	AA969420	Synthetic HAV VP3	367	6	1.1	46	19	AA927955	Staphylococcus aur
295	6	1.1	21	22	AA949691	Modified surfactan	368	6	1.1	46	22	AB935800	SEQ ID NO. 111 fro
296	6	1.1	21	22	AA949691	Modified surfactan	369	6	1.1	46	22	AB935800	SEQ ID NO. 111 fro
297	6	1.1	22	14	AA941151	HAV (type 2) pepti	370	6	1.1	47	23	AA968386	HIV envelope prote
298	6	1.1	22	14	AA941152	HAV (type 2) pepti	371	6	1.1	47	22	AA968386	Human secreted pro
299	6	1.1	22	14	AA941152	Human human diagno	372	6	1.1	47	22	AA968386	Human polypeptide
300	6	1.1	23	22	AA903178	Partial peptide fo	373	6	1.1	48	21	AA956972	Human colon cancer
301	6	1.1	24	16	AA964188	Mouse CD22 domain	374	6	1.1	48	22	AB930808	Peptide #3459 enco
302	6	1.1	25	19	AA950789	Peptide used in im	375	6	1.1	48	22	AB935985	Peptide #3491 enco

376	6	1.1	48	22	ABR21395	Protein #3394 enco	449	6	1.1	60	22	ABR28266	Human peptide #917
377	6	1.1	48	22	AAW56787	Human brain expres	450	6	1.1	60	22	ABR32421	Peptide #5072 enco
378	6	1.1	48	22	AAW69165	Human bone marrow	451	6	1.1	60	22	ABR33442	Peptide #948 enco
379	6	1.1	48	22	AAW16996	Peptide #3430 enco	452	6	1.1	60	22	ABR18901	Protein #900 enco
380	6	1.1	48	22	AAW29488	Peptide #3525 enco	453	6	1.1	60	22	AAW70807	Human bone marrow
381	6	1.1	48	22	AAW4703	Peptide #3385 enco	454	6	1.1	60	22	AAW14493	Peptide #927 enco
382	6	1.1	48	23	ABG38762	Human peptide enco	455	6	1.1	60	22	AAW31112	Peptide #5149 enco
383	6	1.1	49	19	AAW50785	Peptide used in 1m	456	6	1.1	60	23	ABG36273	Human peptide enco
384	6	1.1	49	20	AAW25295	HCV NSS5 carboxy-t	457	6	1.1	60	23	ABG40549	Human peptide enco
385	6	1.1	49	20	AAW25307	HCV NSS5 carboxy-t	458	6	1.1	60	23	ABP05972	Human ORF protein
386	6	1.1	49	20	AAW25319	HCV NSS5 carboxy-t	459	6	1.1	61	21	ABR58216	Lung cancer associ
387	6	1.1	49	20	AAW25330	Protein #569 enco	460	6	1.1	61	22	ABR40897	Peptide #8403 enco
388	6	1.1	49	21	AAW39271	Gene 28 human secr	461	6	1.1	61	22	AAW61757	Human brain expres
389	6	1.1	49	22	ABG16534	Novel human diagno	462	6	1.1	61	22	AAW74553	Human bone marrow
390	6	1.1	49	22	ABR27930	Human peptide #581	463	6	1.1	61	22	AAW44667	Peptide #8704 enco
391	6	1.1	49	22	ABR31242	Peptide #3893 enco	464	6	1.1	61	23	ABG44402	Human peptide enco
392	6	1.1	49	22	ABR33104	Peptide #610 enco	465	6	1.1	62	22	AAU5273	Proionibacterium
393	6	1.1	49	22	ABR36452	Peptide #3958 enco	466	6	1.1	62	22	AAU32279	Novel human secret
394	6	1.1	49	22	ABR18570	Protein #569 enco	467	6	1.1	62	23	ABP01787	Human ORF protein
395	6	1.1	49	22	ABR21795	Protein #3794 enco	468	6	1.1	62	23	ABP11321	Human ORF protein
396	6	1.1	49	22	AAW53901	Human brain expres	469	6	1.1	63	21	AAW62449	Human secreted pro
397	6	1.1	49	22	AAW66289	Human bone marrow	470	6	1.1	63	22	AAW65525	Proionibacterium
398	6	1.1	49	22	AAW69612	Human bone marrow	471	6	1.1	64	18	AAW29898	Mammalian AMPK- $\alpha$ 1p
399	6	1.1	49	22	AAW14157	Peptide #591 enco	472	6	1.1	64	18	AAW29905	Mammalian AMPK- $\alpha$ 1p
400	6	1.1	49	22	AAW17426	Peptide #3860 enco	473	6	1.1	65	22	ABG12137	Novel human diagno
401	6	1.1	49	22	AAW26565	Peptide #602 enco	474	6	1.1	65	22	ABR15793	Human nervous syst
402	6	1.1	49	22	AAW29951	Peptide #3988 enco	475	6	1.1	65	22	ABR10175	Human CDNA SEQ ID
403	6	1.1	49	22	AAW01896	Peptide #578 enco	476	6	1.1	65	23	ABR72318	Rat protein isolat
404	6	1.1	49	22	AAW05109	Peptide #3791 enco	477	6	1.1	66	22	ABG26439	Novel human diagno
405	6	1.1	49	23	ABG35936	Human peptide enco	478	6	1.1	66	22	AAU29965	Novel human secret
406	6	1.1	49	23	ABG39236	Human peptide enco	479	6	1.1	66	22	AAW83291	Human immune/haema
407	6	1.1	50	21	AAW57377	Human peptide enco	480	6	1.1	67	22	AAW81124	Human haematologic
408	6	1.1	50	21	AAW82103	Human haematologic	481	6	1.1	67	23	ABP08057	Human ORF protein
409	6	1.1	51	22	ABR29083	Peptide #1734 enco	482	6	1.1	68	21	AAW12583	Zea mays protein f
410	6	1.1	51	22	ABR34244	Peptide #1750 enco	483	6	1.1	68	22	AAU27808	Bacteriophage Lamb
411	6	1.1	51	22	ABR19680	Protein #1679 enco	484	6	1.1	68	22	AAU27980	Human haematologic
412	6	1.1	51	22	AAW55037	Human brain expres	485	6	1.1	68	22	AAW81075	Human haematologic
413	6	1.1	51	22	AAW67427	Human bone marrow	486	6	1.1	68	22	AAW82046	Human haematologic
414	6	1.1	51	22	AAW15253	Peptide #1687 enco	487	6	1.1	69	22	ABG12136	Novel human diagno
415	6	1.1	51	22	AAW27716	Peptide #1753 enco	488	6	1.1	69	22	ABR28599	Peptide #1250 enco
416	6	1.1	51	22	AAW02997	Peptide #1679 enco	489	6	1.1	69	22	ABR33782	Peptide #1288 enco
417	6	1.1	51	23	ABG37050	Human peptide enco	490	6	1.1	69	22	ABR19225	Protein #1224 enco
418	6	1.1	52	21	AAW27496	Human peptide enco	491	6	1.1	69	22	AAW54548	Human brain expres
419	6	1.1	52	21	AAW56482	Human peptide enco	492	6	1.1	69	22	AAW66955	Human bone marrow
420	6	1.1	52	21	AAW58744	Human peptide enco	493	6	1.1	69	22	AAW14815	Peptide #1249 enco
421	6	1.1	52	22	AAW40304	Proionibacterium	494	6	1.1	69	22	AAW27242	Peptide #1279 enco
422	6	1.1	52	22	AAW31099	Peptide #5136 enco	495	6	1.1	69	22	AAW02540	Peptide #1222 enco
423	6	1.1	53	22	AAW55293	Proionibacterium	496	6	1.1	69	23	ABG36612	Human peptide enco
424	6	1.1	53	23	ABP00140	Human ORF protein	497	6	1.1	70	20	AAU12354	Human 5' EST secre
425	6	1.1	54	22	AAW45763	Proionibacterium	498	6	1.1	70	22	AAU65281	Proionibacterium
426	6	1.1	54	22	ABR29413	Peptide #2064 enco	499	6	1.1	70	22	ABR17923	Human nervous syst
427	6	1.1	54	22	ABR20001	Protein #2000 enco	500	6	1.1	70	22	AAW84635	Human immune/haema
428	6	1.1	54	22	AAW5383	Human brain expres	501	6	1.1	71	20	AAW89021	Polypeptide fragme
429	6	1.1	54	22	AAW67778	Human bone marrow	502	6	1.1	71	21	AAW91558	Human secreted pro
430	6	1.1	54	22	AAW15587	Peptide #2021 enco	503	6	1.1	71	21	AAW91648	Human secreted pro
431	6	1.1	54	22	AAW28079	Peptide #2116 enco	504	6	1.1	71	22	AAU54894	Proionibacterium
432	6	1.1	54	22	AAW03331	Peptide #2013 enco	505	6	1.1	71	22	AAU65864	Proionibacterium
433	6	1.1	54	23	ABG37327	Human peptide enco	506	6	1.1	71	22	ABR51104	Human secreted pro
434	6	1.1	55	22	AAW23748	Human EST enco	507	6	1.1	71	22	AAW89211	Human immune/haema
435	6	1.1	56	22	ABR4174	Peptide #11680 enc	508	6	1.1	72	22	ABG13751	Novel human diagno
436	6	1.1	56	22	AAW65205	Human brain expres	509	6	1.1	73	22	AAU57483	Proionibacterium
437	6	1.1	57	22	AAW21803	Peptide #8237 enco	510	6	1.1	73	23	ABP34183	Human ORF3156 prot
438	6	1.1	57	20	AAW30907	Human secreted pro	511	6	1.1	73	23	ABP08089	Human ORF protein
439	6	1.1	57	22	ABG0982	Novel human diagno	512	6	1.1	74	18	AAW10051	Protein encoded by
440	6	1.1	57	22	AAW03197	Human polypeptide	513	6	1.1	74	22	AAU22955	Novel human enzyme
441	6	1.1	57	23	ABP34431	Human ORF3404 prot	514	6	1.1	74	22	AAW33545	Human colon cancer
442	6	1.1	57	23	ABP07931	Human ORF protein	515	6	1.1	74	22	AAW76595	Human colon cancer
443	6	1.1	59	22	AAW66112	Proionibacterium	516	6	1.1	75	19	AAW77655	Staphylococcus aur
444	6	1.1	59	22	AAW09657	Human polypeptide	517	6	1.1	75	21	AAW19376	Protein encoded by
445	6	1.1	59	22	AAW02704	Human alpha (I) t	518	6	1.1	75	21	AAW70586	Salmonella Pathoge
446	6	1.1	59	22	ABW68058	Amino acid sequenc	519	6	1.1	75	22	AAW83207	Human immune/haema
447	6	1.1	59	23	ABP00864	Human ORF protein	520	6	1.1	76	22	AAW08370	Human polypeptide
448	6	1.1	59	23	ABP04846	Human ORF protein	521	6	1.1	77	20	AAW35865	Chlamydia pneumonia

522	6	1.1	77	21	AA601609	Human secreted pro	595	6	1.1	91	22	AAW73340	Human bone marrow
523	6	1.1	78	16	AA870161	Streptococcus pneu	596	6	1.1	91	22	AA004777	Human polypeptide
524	6	1.1	78	21	AA612582	Kea may's protein f	597	6	1.1	91	22	AAAI5227	Peptide #1661 enco
525	6	1.1	78	22	AA064153	Proionibacterium	598	6	1.1	91	22	AAAI9921	Peptide #6355 enco
526	6	1.1	79	21	AA034413	Human secreted pro	599	6	1.1	91	22	AAW27688	Peptide #1725 enco
527	6	1.1	79	22	AA051485	Proionibacterium	600	6	1.1	91	22	AAW33542	Peptide #7579 enco
528	6	1.1	79	22	AB030529	Peptide #3180 enco	601	6	1.1	91	22	AAW02970	Peptide #1652 enco
529	6	1.1	79	22	AB035650	Peptide #3196 enco	602	6	1.1	91	23	AB037023	Human peptide enco
530	6	1.1	79	22	AB021122	Protein #3121 enco	603	6	1.1	91	23	AB043193	Human peptide enco
531	6	1.1	79	22	AAW65055	Human brain expres	604	6	1.1	92	21	AAW63220	Human secreted pro
532	6	1.1	79	22	AAW68885	Human bone marrow	605	6	1.1	92	21	AAW53378	Arabisdopsis thalia
533	6	1.1	79	22	AAW16708	Peptide #3142 enco	606	6	1.1	92	22	AAW62580	Proionibacterium
534	6	1.1	79	22	AAW29195	Peptide #3232 enco	607	6	1.1	92	22	AAW07902	Human polypeptide
535	6	1.1	79	22	AAW04424	Peptide #3106 enco	608	6	1.1	92	22	AAW72521	Human polypeptide
536	6	1.1	79	22	AAW0426	Gene #7 associated	609	6	1.1	92	23	ABW80905	Murine MCP3 fragme
537	6	1.1	79	23	AB038465	Human peptide enco	610	6	1.1	92	23	AAW80162	Human motor protei
538	6	1.1	80	21	AAW03669	Human secreted pro	611	6	1.1	93	21	AAW05283	Arabisdopsis thalia
539	6	1.1	80	22	AAW65636	Human immune/haema	612	6	1.1	93	21	AAW27495	Arabisdopsis thalia
540	6	1.1	80	22	AAW65986	Human immune/haema	613	6	1.1	93	21	AAW45310	Arabisdopsis thalia
541	6	1.1	80	23	ABW05418	Human OREF protein	614	6	1.1	93	21	AAW45319	Arabisdopsis thalia
542	6	1.1	81	11	AAW04690	Fusion of prochymo	615	6	1.1	94	16	AAW67389	C. perfingens Groe
543	6	1.1	81	22	AAW02978	Human polypeptide	616	6	1.1	94	21	AAW37653	Arabisdopsis thalia
544	6	1.1	82	20	AAW85172	Murine pituitary-d	617	6	1.1	94	22	ABW09576	Novel human diagno
545	6	1.1	82	21	AAW33971	Arabisdopsis thalia	618	6	1.1	94	22	AAW89143	Human secreted pro
546	6	1.1	82	22	ABW08251	Novel human diagno	619	6	1.1	95	22	AAW3047	Proionibacterium
547	6	1.1	82	23	AAW80941	Human transcriptio	620	6	1.1	95	22	AAW30319	C glutamylum prote
548	6	1.1	82	23	AAW79431	Human transcriptio	621	6	1.1	96	21	AAW55377	Arabisdopsis thalia
549	6	1.1	83	18	AAW31383	Rat type G protein	622	6	1.1	96	22	AAW45307	Proionibacterium
550	6	1.1	83	19	AAW77650	Staphylococcus aur	623	6	1.1	96	22	ABW04785	Novel human diagno
551	6	1.1	83	20	AAW01249	Chemokine receptor	624	6	1.1	96	22	ABW36220	Novel human diagno
552	6	1.1	83	20	AAW97225	Rat type 1 ligand po	625	6	1.1	96	22	AAW23933	Sheep EST encoded
553	6	1.1	83	21	AAW10354	Rat oxytocin secre	626	6	1.1	97	16	AAW70803	Growth factor-acti
554	6	1.1	83	21	AAW11943	Arabisdopsis thalia	627	6	1.1	97	16	AAW69995	Revised complete r
555	6	1.1	83	22	AAW62523	Rat CRH releasing	628	6	1.1	97	22	AAW15667	Human nervous syst
556	6	1.1	83	22	AAW61170	Anopheles albimanu	629	6	1.1	98	17	AAW30087	Human chemokine be
557	6	1.1	84	22	AAW64348	Proionibacterium	630	6	1.1	98	18	AAW30191	Monocyte chemotact
558	6	1.1	84	23	ABW06052	Human OREF protein	631	6	1.1	98	18	AAW31382	Bovine genome derl
559	6	1.1	85	17	AAW04170	Flca calreticulin	632	6	1.1	98	18	AAW31368	Bovine G protein-c
560	6	1.1	85	21	AAW24937	Arabisdopsis thalia	633	6	1.1	98	18	AAW22670	Human chemokine be
561	6	1.1	85	21	AAW96827	N. thaliana PLP3 c	634	6	1.1	98	19	AAW56087	Human monocyte che
562	6	1.1	86	11	AAW06502	GST-50 clone-enco	635	6	1.1	98	20	AAW41164	Human chemokine be
563	6	1.1	86	22	AAW04259	Human gene 6 enco	636	6	1.1	98	20	AAW97224	Bovine genome-driv
564	6	1.1	87	15	AAW9654	KCV peptide C14-2-	637	6	1.1	98	20	AAW97217	Bovine pituitary-d
565	6	1.1	87	22	AAW36170	Klebsiella pneumon	638	6	1.1	98	20	AAW95187	Bovine genome-derl
566	6	1.1	87	22	AAW54569	Proionibacterium	639	6	1.1	98	21	AAW15808	Human chemokine CC
567	6	1.1	87	22	AAW30956	Novel human secret	640	6	1.1	98	21	AAW15831	Human chemokine MI
568	6	1.1	87	23	AAW88701	Human immune/haema	641	6	1.1	98	21	AAW10346	Bovine oxytocin se
569	6	1.1	87	23	AAW99920	Human 49875 helica	642	6	1.1	98	21	AAW10353	Bovine oxytocin se
570	6	1.1	88	19	AAW38753	Streptococcus pneu	643	6	1.1	98	21	AAW23146	Arabisdopsis thalia
571	6	1.1	88	21	AAW07191	Arabisdopsis thalia	644	6	1.1	98	21	AAW40005	Arabisdopsis thalia
572	6	1.1	88	21	AAW57411	Arabisdopsis thalia	645	6	1.1	98	21	AAW60540	Arabisdopsis thalia
573	6	1.1	88	23	ABW04722	Human OREF protein	646	6	1.1	98	21	AAW95534	Human chemokine be
574	6	1.1	89	21	AAW24936	Arabisdopsis thalia	647	6	1.1	98	22	AAW83415	Human ion2a protei
575	6	1.1	89	21	AAW49862	Arabisdopsis thalia	648	6	1.1	98	22	AAW62515	Bovine CRH releasi
576	6	1.1	89	21	AAW57197	Arabisdopsis thalia	649	6	1.1	98	22	AAW62522	Bovine CRH releasi
577	6	1.1	89	22	AAW34371	Staphylococcus aur	650	6	1.1	98	22	AAW31795	Amino acid sequenc
578	6	1.1	89	22	ABW30189	Peptide #2840 enco	651	6	1.1	98	23	AAW77180	Human chemokine MC
579	6	1.1	89	22	ABW35354	Peptide #2860 enco	652	6	1.1	98	23	ABW04626	Human OREF protein
580	6	1.1	89	22	ABW20794	Protein #2793 enco	653	6	1.1	98	23	AAW01997	Protein of human c
581	6	1.1	89	22	AAW56185	Human brain expres	654	6	1.1	98	23	ABW55286	Lactococcus lactis
582	6	1.1	89	22	AAW68558	Human bone marrow	655	6	1.1	98	23	AAW15751	Human chemokine be
583	6	1.1	89	22	AAW16366	Peptide #2800 enco	656	6	1.1	98	23	AAW95387	Human monocyte che
584	6	1.1	89	22	AAW28863	Peptide #2900 enco	657	6	1.1	99	11	AAW63398	Human MCF precurs
585	6	1.1	89	22	AAW04104	Peptide #2786 enco	658	6	1.1	99	13	AAW26581	Sequence of P6 pre
586	6	1.1	89	23	ABW38136	Human peptide enco	659	6	1.1	99	13	AAW28663	MCF. Synthetic.
587	6	1.1	90	20	AAW11942	Human 5' EST secre	660	6	1.1	99	16	AAW73914	Human monocyte che
588	6	1.1	90	23	ABW31460	Human OREF prote	661	6	1.1	99	16	AAW70801	Chemottractant pr
589	6	1.1	91	22	ABW29056	Peptide #1707 enco	662	6	1.1	99	16	AAW70800	Chemottractant pr
590	6	1.1	91	22	ABW39923	Peptide #7429 enco	663	6	1.1	99	19	AAW40174	Macrophage chemoa
591	6	1.1	91	22	ABW19653	Protein #1652 enco	664	6	1.1	99	20	AAW48391	Human prostate can
592	6	1.1	91	22	ABW24474	Protein #6473 enco	665	6	1.1	99	20	AAW26176	Monocyte chemotact
593	6	1.1	91	22	AAW55011	Human brain expres	666	6	1.1	99	20	AAW05300	C-C chemokine, MCP
594	6	1.1	91	22	AAW60668	Human brain expres	667	6	1.1	99	20	AAW07233	Wild type monocyte

668	6	1.1	99	20	AAV07237	Wild type monocyte	741	6	1.1	114	20	AAV12363	Human 5' EST secre
669	6	1.1	99	21	AAAB15785	Human chemokine MC	742	6	1.1	114	22	AAE12624	Human gene 2 encod
670	6	1.1	99	21	AAAB15787	Human chemokine MC	743	6	1.1	114	22	AAU04354	Mammalian toxicolo
671	6	1.1	99	21	AAAG40924	Zea mays protein f	744	6	1.1	114	22	AAAG3282	Human protein HP10
672	6	1.1	99	22	AAAB6546	Human testicular a	745	6	1.1	114	23	AAAB89137	Human polypeptide
673	6	1.1	99	22	AAAB6542	Human reproductiv	746	6	1.1	115	22	AAU48455	Protonibacterium
674	6	1.1	99	22	AAAB7914	Human monocyte che	747	6	1.1	115	22	ABG03093	Novel human diagno
675	6	1.1	99	23	AAAB80901	Human MCP1 fragmen	748	6	1.1	115	22	ABAB2998	Peptide #2649 enco
676	6	1.1	99	23	AAAM52440	HIV_Nef1 fusion pr	749	6	1.1	115	22	ABAB20604	Protein #2603 enco
677	6	1.1	99	23	AAU077179	Human small induc	750	6	1.1	115	22	AAAB56002	Human brain expres
678	6	1.1	99	23	AAAO19999	99-mer protein of	751	6	1.1	115	22	AAAB68371	Human bone marrow
679	6	1.1	100	20	AAAB88231	HIV-1 co-receptor	752	6	1.1	115	22	AAAB6188	Peptide #2622 enco
680	6	1.1	100	21	AAAG02737	Human secreted pro	753	6	1.1	115	22	AAAB24360	Human EST encoded
681	6	1.1	100	22	AAU64365	Protonibacterium	754	6	1.1	115	22	AAAM03918	Peptide #2600 enco
682	6	1.1	100	22	AAE23969	Human TIP39 precu	755	6	1.1	115	23	ABG37934	Human peptide enco
683	6	1.1	101	21	ABAB49169	Listeria monocytog	756	6	1.1	116	21	AAV75943	Rat skin cell secr
684	6	1.1	101	21	AAAG27652	Arabidopsis thalia	757	6	1.1	116	21	AAAB55882	Skin cell protein,
685	6	1.1	101	21	AAAG33970	Arabidopsis thalia	758	6	1.1	116	23	ABAB2082	Rat protein isolat
686	6	1.1	101	22	AAAG47102	Protonibacterium	759	6	1.1	117	19	ABAB27168	Cooperia oncophora
687	6	1.1	101	22	ABG13749	Novel human diagno	760	6	1.1	117	22	ABAB28381	Peptide #1032 enco
688	6	1.1	101	22	AAAO0287	Human polypeptide	761	6	1.1	117	22	ABAB33559	Peptide #1065 enco
689	6	1.1	101	22	AAEO2705	Human alpha (I) t	762	6	1.1	117	22	ABAB19016	Human peptide enco
690	6	1.1	101	22	ABAB6059	Amino acid sequenc	763	6	1.1	117	22	AAAB54333	Human brain expres
691	6	1.1	101	22	ABP00815	Human OREX protein	764	6	1.1	117	22	AAAB6733	Human bone marrow
692	6	1.1	102	20	AAV11993	Human 5' EST secre	765	6	1.1	117	22	AAAB4559	Peptide #1033 enco
693	6	1.1	102	22	AAAB7970	Human immune/haema	766	6	1.1	117	22	AAAB27020	Peptide #1057 enco
694	6	1.1	103	13	AAAR29890	HCV NS4-NS5 peptid	767	6	1.1	117	22	AAAM02324	Peptide #1006 enco
695	6	1.1	103	21	AAAG24935	Arabidopsis thalia	768	6	1.1	117	23	ABG36387	Human peptide enco
696	6	1.1	103	21	AAAG49861	Arabidopsis thalia	769	6	1.1	117	23	ABP10316	Human OREX protein
697	6	1.1	103	22	AAU59174	Protonibacterium	770	6	1.1	118	17	AAAM05521	HCMV Towne strain
698	6	1.1	104	19	AAAB6088	Novel human diagno	771	6	1.1	118	22	AAU41271	Protonibacterium
699	6	1.1	104	19	AAAB57322	Murine monocyte ch	772	6	1.1	118	22	ABG05920	Novel human diagno
700	6	1.1	104	19	AAU41693	Protonibacterium	773	6	1.1	118	22	ABG10573	Novel human diagno
701	6	1.1	104	22	AAAB3754	Human polypeptide,	774	6	1.1	118	22	ABAB22427	Novel human diagno
702	6	1.1	105	18	AAAB27978	Amino acid sequenc	775	6	1.1	118	22	ABAB34892	Peptide #2371 enco
703	6	1.1	105	23	ABG05145	Novel human diagno	776	6	1.1	118	22	ABAB29720	Peptide #2398 enco
704	6	1.1	105	23	ABG33152	Human ORE2125 prot	777	6	1.1	118	22	ABAB20309	Protein #2308 enco
705	6	1.1	106	22	ABG13572	Novel human diagno	778	6	1.1	118	22	AAU32460	Novel human secre
706	6	1.1	106	22	ABG18322	Novel human diagno	779	6	1.1	118	22	AAAB5702	Human brain expres
707	6	1.1	106	22	ABG18322	Novel human diagno	780	6	1.1	118	22	AAAB6080	Human bone marrow
708	6	1.1	107	13	AAAR24275	NANBH specific ant	781	6	1.1	118	22	AAAB15905	Peptide #2339 enco
709	6	1.1	108	15	AAAR51502	Pilin protein vari	782	6	1.1	118	22	AAAM28410	Peptide #2447 enco
710	6	1.1	108	15	AAAR51503	Pilin protein vari	783	6	1.1	118	22	AAAM03640	Peptide #2322 enco
711	6	1.1	108	22	AAO13271	Human polypeptide	784	6	1.1	118	23	ABAB1433	Human prostate spe
712	6	1.1	109	13	AAAR24353	Cytokine encoded b	785	6	1.1	118	23	ABG37607	Human prostate enco
713	6	1.1	109	18	AAAR26655	Human, beta-chemok	786	6	1.1	118	23	ABAP41730	Human ovarian anti
714	6	1.1	109	19	AAAG2072	Human MC proprotel	787	6	1.1	119	21	ABAB33033	Pinus radiata tran
715	6	1.1	109	21	AAAG02147	Human secreted pro	788	6	1.1	119	22	ABAB71029	Drosophila melanog
716	6	1.1	109	22	AAU50156	Protonibacterium	789	6	1.1	119	22	ABG16892	Novel human diagno
717	6	1.1	109	22	AAU27814	Full-length polype	790	6	1.1	120	21	AAAB42520	Human OREX ORE2284
718	6	1.1	109	23	AAO21490	Shell inducible cy	791	6	1.1	120	22	ABAB69393	Drosophila melanog
719	6	1.1	109	23	AAAB52441	HIV_Nef1 fusion pr	792	6	1.1	120	22	AAU43965	Protonibacterium
720	6	1.1	109	23	ABG35167	Reference sequence	793	6	1.1	120	22	AAU49944	Protonibacterium
721	6	1.1	109	23	ABG35168	Polymorphic varian	794	6	1.1	120	22	ABG27167	Novel human diagno
722	6	1.1	109	23	ABG35169	Polymorphic varian	795	6	1.1	120	22	AAU27942	Human contig polyp
723	6	1.1	109	23	ABP04061	Human OREX protein	796	6	1.1	120	22	AAO01982	Human polypeptide
724	6	1.1	110	20	AAAB67811	Human secreted pro	797	6	1.1	120	22	ABAB3759	Lactococcus lactis
725	6	1.1	110	21	ABAB32405	Human secreted pro	798	6	1.1	121	21	AAAB53759	Human prostate can
726	6	1.1	110	21	AAAG56507	Arabidopsis thalia	799	6	1.1	121	21	AAU08295	Human PML-like pro
727	6	1.1	110	22	ABG12594	Novel human diagno	800	6	1.1	121	22	AAU27435	Novel bone marrow
728	6	1.1	110	23	ABAB80904	Human MCP3 fragmen	801	6	1.1	121	22	AAO09604	Human polypeptide
729	6	1.1	110	23	ABAB8265	Human MCP3 fragmen	802	6	1.1	121	22	AAAB63745	Human prostate can
730	6	1.1	110	23	ABP02700	Human OREX protein	803	6	1.1	122	22	AAO05928	Human prostate can
731	6	1.1	112	23	ABAB35094	Human ORE4067 prot	804	6	1.1	123	18	AAAB28120	Human polypeptide
732	6	1.1	112	18	AAAB10566	Human aortic pref	805	6	1.1	123	21	AAAG45318	Amino acid sequenc
733	6	1.1	113	18	AAAM10565	Rat aortic prefere	806	6	1.1	123	21	ABG23572	Novel human diagno
734	6	1.1	113	19	AAAM77046	Rat aortic prefere	807	6	1.1	123	22	AAAB2992	Human immune/haema
735	6	1.1	113	19	AAAM77047	Human aortic-prefe	808	6	1.1	123	22	AAAB35108	Synechocystis clps
736	6	1.1	113	21	AAAG12422	Zea mays protein f	809	6	1.1	123	22	ABP01786	Human OREX protein
737	6	1.1	113	21	AAV70076	Rat aortic-prefere	810	6	1.1	124	23	AAAG75577	Human colon cancer
738	6	1.1	113	21	AAV70077	Human aortic-prefe	811	6	1.1	125	14	AAAR44405	NANBH derived pol
739	6	1.1	113	21	AAV70080	Mouse aortic-prefe	812	6	1.1	125	14	AAAR44407	NANBH derived pol
740	6	1.1	113	21	AAV70081	Aortic-prefere	813	6	1.1	125	21	AAAG01678	Human secreted pro

814	6	1.1	125	22	ABG17349	Novel human diagno
815	6	1.1	125	22	ABG22420	Novel human diagno
816	6	1.1	125	23	AAAG80303	Human cellular dif
817	6	1.1	125	23	AAAG7261	Human cellular dif
818	6	1.1	126	16	AAAG00640	Infectious laryngo
819	6	1.1	126	17	AAAG06792	IL1V unique short
820	6	1.1	126	21	AAAG27864	Protein fragment e
821	6	1.1	126	21	AAAG37443	Staphylococcus aur
822	6	1.1	126	22	AAAG00187	Human polypeptide
823	6	1.1	127	20	AAAG12339	Human 5' EST seque
824	6	1.1	127	22	ABG27278	Novel human diagno
825	6	1.1	127	22	ABG03589	Human musculoskele
826	6	1.1	127	22	AAAG3182	Human polypeptide,
827	6	1.1	128	20	AAAG29538	Human lung tumour
828	6	1.1	128	21	AAAG44443	Human lung tumour
829	6	1.1	128	22	AAAG13784	Human lung tumour
830	6	1.1	128	22	AAAG04223	Human gene 6 encod
831	6	1.1	128	23	ABG64481	Human albumin fusl
832	6	1.1	129	21	AAAG3866	Heavy chain (VH) g
833	6	1.1	129	21	AAAG4836	Protonibacterium
834	6	1.1	129	22	ABG00569	Novel human diagno
835	6	1.1	129	22	ABP29700	Staphylococcus epi
836	6	1.1	130	22	AAAG1625	Protonibacterium
837	6	1.1	130	22	AAAG5002	Protonibacterium
838	6	1.1	130	22	ABG06340	Novel human diagno
839	6	1.1	131	21	AAAG56506	Aribidopsis thalia
840	6	1.1	131	21	ABG69514	Drosophila melanog
841	6	1.1	131	22	AAAG1033	Protonibacterium
842	6	1.1	131	23	ABAG49927	Histocerta monocyto
843	6	1.1	132	21	AAAG4405	Aribidopsis thalia
844	6	1.1	132	21	AAAG3065	Protonibacterium
845	6	1.1	133	21	AAAG19326	Amino acid sequenc
846	6	1.1	133	21	ABG5943	Human secreted pro
847	6	1.1	133	23	ABG72388	Murine protein iso
848	6	1.1	134	20	AAAG89023	polypeptide framge
849	6	1.1	134	20	AAAG89023	Sequence ID #723 f
850	6	1.1	134	21	AAAG19317	Amino acid sequenc
851	6	1.1	134	21	AAAG19323	Amino acid sequenc
852	6	1.1	134	21	AAAG19324	Amino acid sequenc
853	6	1.1	134	21	AAAG05282	Aribidopsis thalia
854	6	1.1	134	21	AAAG30299	Aribidopsis thalia
855	6	1.1	134	22	ABG51103	Human secreted pro
856	6	1.1	134	22	ABG51106	Human secreted pro
857	6	1.1	134	22	AAAG32534	Novel human secret
858	6	1.1	134	22	AAAG4311	Human reproductive
859	6	1.1	134	22	AAAG04260	Human gene 6 encod
860	6	1.1	135	19	AAAG5058	Lettuce antifungal
861	6	1.1	135	21	AAAG4098	Human cancer assoc
862	6	1.1	135	22	AAAG30647	C glutamincum prote
863	6	1.1	136	20	AAAG8593	Secreted protein e
864	6	1.1	136	22	ABG00106	Novel human diagno
865	6	1.1	136	22	ABG50517	Human secreted pro
866	6	1.1	137	21	AAAG45309	Aribidopsis thalia
867	6	1.1	137	22	ABG50380	Human secreted pro
868	6	1.1	137	22	ABG68578	Murine MOC1 pepti
869	6	1.1	137	23	ABG83156	Partial OMPA prote
870	6	1.1	137	23	AAAG0552	Human g protein-co
871	6	1.1	138	22	ABG25114	Novel human diagno
872	6	1.1	138	22	AAAG27770	Human full-length
873	6	1.1	138	22	AAAG68495	Amino acid sequenc
874	6	1.1	139	20	AAAG12341	Human 5' EST seque
875	6	1.1	139	21	AAAG58965	Breast and ovarian
876	6	1.1	139	21	AAAG43865	Heavy chain (VH) g
877	6	1.1	140	14	AAAG51076	Polypeptide encode
878	6	1.1	140	14	AAAG33784	Aribidopsis thalia
879	6	1.1	140	21	AAAG43872	Heavy chain (VH) g
880	6	1.1	140	21	AAAG20380	Soybean sterol del
881	6	1.1	141	20	AAAG85737	Polypeptide with t
882	6	1.1	141	21	AAAG43859	Heavy chain framew
883	6	1.1	141	21	AAAG43869	Heavy chain (VH) g
884	6	1.1	141	21	AAAG43873	Heavy chain (VH) g
885	6	1.1	141	21	ABG64190	Drosophila melanog
886	6	1.1	141	22	AAAG6017	Human immune/haema
887	6	1.1	141	22	AAAG92123	C glutamincum prote
888	6	1.1	141	22	AAAG8353	Human secreted pro
889	6	1.1	141	22	ABP42622	Human ovarian anti
890	6	1.1	142	18	AAAG20352	H. pylori inner me
891	6	1.1	142	20	AAAG37322	Amino acid sequenc
892	6	1.1	142	21	AAAG3858	Heavy chain framew
893	6	1.1	142	21	AAAG3868	Heavy chain (VH) g
894	6	1.1	142	21	AAAG3871	Heavy chain (VH) g
895	6	1.1	142	21	ABG09219	Novel human diagno
896	6	1.1	143	21	AAAG3867	Heavy chain (VH) g
897	6	1.1	143	22	AAAG47041	Protonibacterium
898	6	1.1	143	22	AAAG12965	Human polypeptide
899	6	1.1	143	22	AAAG13481	Human polypeptide
900	6	1.1	144	21	AAAG25420	Pinus radiata cell
901	6	1.1	144	21	AAAG49860	Aribidopsis thalia
902	6	1.1	144	21	AAAG9884	Human haemopoietin
903	6	1.1	144	21	AAAG9887	Mouse haemopoietin
904	6	1.1	144	21	AAAG4933	Protonibacterium
905	6	1.1	144	22	ABP02955	Human ORFX protein
906	6	1.1	144	23	ABG55182	Lactococcus lactis
907	6	1.1	145	11	AAAG05586	Acetohydroxy acid
908	6	1.1	145	11	AAAG4097	Human cancer assoc
909	6	1.1	145	23	ABP34941	Human ORF3914 prot
910	6	1.1	146	11	AAAG06500	GST-42 clone encod
911	6	1.1	146	19	AAAG71073	MSRV-1 clone 41 po
912	6	1.1	146	19	AAAG71074	MSRV-1 clone 42 po
913	6	1.1	146	19	AAAG71075	MSRV-1 clone 43 po
914	6	1.1	147	19	AAAG8282	H. pylori GHPO 114
915	6	1.1	147	21	AAAG3861	Heavy chain framew
916	6	1.1	147	21	AAAG29423	Human g protein-co
917	6	1.1	147	23	ABG60711	Novel g protein co
918	6	1.1	148	21	AAAG16471	Aribidopsis thalia
919	6	1.1	148	21	AAAG30298	Aribidopsis thalia
920	6	1.1	148	21	AAAG33783	Aribidopsis thalia
921	6	1.1	148	21	AAAG50342	Aribidopsis thalia
922	6	1.1	148	21	ABG14195	Novel human diagno
923	6	1.1	149	22	AAAG34867	E. coli cellular p
924	6	1.1	149	22	AAAG38267	Salmonella typhi c
925	6	1.1	151	22	ABG68382	Drosophila melanog
926	6	1.1	151	22	AAAG5916	Human protein sequ
927	6	1.1	151	22	AAAG42142	Human polypeptide
928	6	1.1	152	22	AAAG4927	Human protein sequ
929	6	1.1	153	21	AAAG41527	Human ORFX ORF1291
930	6	1.1	153	21	AAAG0453	Protonibacterium
931	6	1.1	153	22	AAAG7641	Human immune/haema
932	6	1.1	153	22	AAAG11056	Human polypeptide
933	6	1.1	153	23	ABP00805	Human ORFX protein
934	6	1.1	154	19	AAAG72336	Pathogen response
935	6	1.1	154	22	ABG00333	Novel human diagno
936	6	1.1	154	23	ABG54224	Lactococcus lactis
937	6	1.1	155	23	ABP02097	Human ORFX protein
938	6	1.1	156	20	AAAG29907	Murine MCP-3 and h
939	6	1.1	156	21	AAAG06747	Aribidopsis thalia
940	6	1.1	156	21	AAAG40008	Aribidopsis thalia
941	6	1.1	157	19	AAAG54444	Mouse novel secret
942	6	1.1	157	20	AAAG37718	Amino acid sequenc
943	6	1.1	157	21	AAAG10251	Human adult testes
944	6	1.1	158	21	AAAG69891	Human NR8beta/FLAG
945	6	1.1	158	23	ABP25702	Streptococcus poly
946	6	1.1	159	20	AAAG75355	Chlamydia trachoma
947	6	1.1	159	21	AAAG45308	Aribidopsis thalia
948	6	1.1	159	21	ABG66311	Drosophila melanog
949	6	1.1	160	21	AAAG59429	Wheat inositol 1,3
950	6	1.1	160	22	ABG00753	Novel human diagno
951	6	1.1	160	22	ABG52755	Escherichia coli p
952	6	1.1	161	21	AAAG49838	Protonibacterium
953	6	1.1	161	22	AAAG49030	Protonibacterium
954	6	1.1	161	22	AAAG4850	Novel human diagno
955	6	1.1	162	22	ABG5181	Human polypeptide
956	6	1.1	162	23	ABG89143	Human polypeptide
957	6	1.1	163	12	AAAG5060	Murine anti-ICAM m
958	6	1.1	163	12	AAAG5200	R6-5-D6 anti-ICAM-
959	6	1.1	163	20	AAAG59861	Human normal uteru

960	6	1.1	163	21	AA619624	Arbidiopsis thalia
961	6	1.1	163	22	ABG06103	Novel human diago
962	6	1.1	164	21	AA619412	Arbidiopsis thalia
963	6	1.1	164	22	ABG96013	Human testicular a
964	6	1.1	164	22	AA653918	Protonibacterium
965	6	1.1	164	22	AA695317	Human reproductiv
966	6	1.1	164	22	ABR49408	Listeria monocytog
967	6	1.1	165	22	AA002533	Human polypeptide
968	6	1.1	166	21	AA625307	Escherichia grandis
969	6	1.1	166	22	ABG05346	Novel human diago
970	6	1.1	167	21	AA623145	Arbidiopsis thalia
971	6	1.1	167	21	AA605539	Arbidiopsis thalia
972	6	1.1	167	22	ABG69928	Drosophila melanog
973	6	1.1	168	20	AA634973	Chlamydia pneumoni
974	6	1.1	168	22	ABG60047	Drosophila melanog
975	6	1.1	171	20	AA629909	Human MCP-3 and H
976	6	1.1	171	21	AA640004	Arbidiopsis thalia
977	6	1.1	171	21	AA676225	Fragment of human
978	6	1.1	171	22	ABG10768	Novel human diago
979	6	1.1	171	22	ABG90289	C. glutamicum prote
980	6	1.1	172	18	AA655500	H. pylori ORF 019e
981	6	1.1	172	22	AA655245	Human protein sequ
982	6	1.1	173	18	AA655388	H. pylori ORF 119e
983	6	1.1	173	21	AA614772	Arbidiopsis thalia
984	6	1.1	173	22	ABG26625	Novel human diago
985	6	1.1	173	23	ABP30456	Streptococcus poly
986	6	1.1	174	19	AA638627	Streptococcus pneu
987	6	1.1	175	13	AA620565	Non-A non-B hepati
988	6	1.1	175	21	AA634282	Human secreted pro
989	6	1.1	175	22	ABG01270	Novel human diago
990	6	1.1	175	22	AA620500	Human secreted pro
991	6	1.1	175	23	AA611823	Human secreted pro
992	6	1.1	176	21	ABG56969	Human prostate can
993	6	1.1	176	22	ABP70814	Drosophila melanog
994	6	1.1	176	22	AA635246	Enterococcus faeca
995	6	1.1	177	21	AA658892	Breast and ovarian
996	6	1.1	177	21	AA616470	Arbidiopsis thalia
997	6	1.1	177	21	AA619623	Arbidiopsis thalia
998	6	1.1	177	21	AA600869	Human secreted pro
999	6	1.1	177	21	AA600869	Human secreted pro
1000	6	1.1	177	22	AB611068	Human cation chan

## ALIGNMENTS

## RESULT 1

ID AAY75753 standard; Protein: 508 AA.

XX	AAV75753;	
AC	AAV75753;	
XX		
DT	21-MAR-2000 (first entry)	
XX		
DE	Neisseria meningitidis ORF 987 protein sequence SEQ ID NO:2978.	
XX		
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;	
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;	
KW	antibacterial; gene therapy.	
XX		
OS	Neisseria meningitidis.	
XX		
PN	W09957280-A2.	
XX		
PD	11-NOV-1999.	
XX		
PF	30-APR-1999; 99WO-US09346.	
XX		
PR	01-MAY-1998; 98US-0083758.	
PR	31-JUL-1998; 98US-0094869.	
PR	02-SEP-1998; 98US-0098994.	
PR	02-SEP-1998; 98US-0099062.	
PR	09-OCT-1998; 98US-0103749.	

PR	09-OCT-1998; 98US-0103794.	
PR	09-OCT-1998; 98US-0103796.	
PR	25-FEB-1999; 98US-0121528.	
XX		
FA	(CHIR) CHIRON CORP.	
XX	(GENO-) INST GENOMIC RES.	
XX		
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;	
PI	Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;	
PI	Tetzelin H, Venter JC.	
XX		
DR	WPI: 2000-062150/05.	
DR	N-PSDB: AA254515.	
XX		
PT	Novel Neisserial polypeptides predicted to be useful antigens for	
PT	vaccines and diagnostics	
XX		
PS	Claim 2; Page 1396-1397; 1453pp; English.	
XX		
CC	AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941	
CC	represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides	
CC	and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent	
CC	PCR primers used in the exemplification of the present invention. The	
CC	polypeptides, the polynucleotides, antibodies and compositions of	
CC	the invention can be used as vaccines, as diagnostic reagents, and as	
CC	immunogenic compositions. The polypeptides can be used in the	
CC	manufacture of medicaments for treating or preventing infection due to	
CC	Neisseria bacteria (e.g. meningitis and septicemia), to detect the	
CC	presence of Neisseria bacteria, or to raise antibodies. They may also	
CC	be used to screen for agonists or antagonists, which may themselves	
CC	may also be used in gene therapy protocols.	
XX		
SO	Sequence 508 AA;	
	Query Match 96.8%; Score 508; DB 21; Length 508;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 508; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	18 MKTRSLISLCLLCSCGSMPLPLEERTESRHNFTSKPRVDNIIQIRHPHTNGLSDIY 77	
DB	1 MKTRSLISLCLLCSCGSMPLPLEERTESRHNFTSKPRVDNIIQIRHPHTNGLSDIY 60	
QY	78 LNDPHEAFARALITSAEHSIDLOYIWRNDISGRLEFVLYLAERGVRLLLDDN 137	
DB	61 LNDPHEAFARALITSAEHSIDLOYIWRNDISGRLEFVLYLAERGVRLLLDDN 120	
QY	138 NTRGLDILLALDSHPNENVRLEFNPFLRKWRALGYLTDFPRLNRRMHNKSFADNRATI 197	
DB	121 NTRGLDILLALDSHPNENVRLEFNPFLRKWRALGYLTDFPRLNRRMHNKSFADNRATI 180	
QY	198 LGGRNIGDEYFKVGEDVYFADLLATGSGVYGEVSHDFDRYWASHAHMNRITIRSGNIG 257	
DB	181 LGGRNIGDEYFKVGEDVYFADLLATGSGVYGEVSHDFDRYWASHAHMNRITIRSGNIG 240	
QY	258 KGLALGYNDETSRHALRRETYEVSPLYOKIOTGRIDMOSVOTRLISDDPAAGLDRDR 317	
DB	241 KGLALGYNDETSRHALRRETYEVSPLYOKIOTGRIDMOSVOTRLISDDPAAGLDRDR 300	
QY	318 RKPPIAGRLDALKQPEKSVYLVSPYFVPTKSGDALAKLVQSDIDVTVLNSIQATDVA 377	
DB	301 RKPPIAGRLDALKQPEKSVYLVSPYFVPTKSGDALAKLVQSDIDVTVLNSIQATDVA 360	
QY	378 AVHSGVYKRRPPLKAGIKIYELQPNNAVPATKDKGLTSSVTSIHAKTFLVDGKRIFIG 437	
DB	361 AVHSGVYKRRPPLKAGIKIYELQPNNAVPATKDKGLTSSVTSIHAKTFLVDGKRIFIG 420	
QY	438 SFNIDPSARLNTMGVYIESPKIAEQMERTLATSEYAYRVLDRNRLQWMDPATRK 497	
DB	421 SFNIDPSARLNTMGVYIESPKIAEQMERTLATSEYAYRVLDRNRLQWMDPATRK 480	
QY	498 TYPNEPEAKLWKRIAAKILSLPTESLL 525	

DB 481 TYPNEPEAKLMKRIAAILSLPIESLL 508

RESULT 2

AAV75752

ID AAV75752 standard; Protein; 508 AA.

XX AAV75752;

DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 987 protein sequence SEQ ID NO:2976.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.

XX Neisseria meningitidis.

OS Neisseria meningitidis.

XX WO957280-A2.

XX 11-NOV-1999.

PD 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

DR WPI: 2000-062150/05.

DR N-PSDB: AAZ54514.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2: Page 1395; 1453pp; English.

PS AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941

CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

XX Sequence 508 AA;

SO Query Match 45.1%; Score 237; DB 21; Length 508;

Best Local Similarity 100.0%; Pred. No. 2.7e-226;

Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 MKTSLSLILCLLCCSSMLPPEERTESRHFNTSKVRLLDNLQIRHPIHNGLSIDY 77

DB 1 MKTSLSLILCLLCCSSMLPPEERTESRHFNTSKVRLLDNLQIRHPIHNGLSIDY 60

OY 78 LNDPHEAFARAALISAEHSLDLOYIWMNDISGRLLFNVLAAEGVRRLLDDN 137

DB 61 LNDPHEAFARAALISAEHSLDLOYIWMNDISGRLLFNVLAAEGVRRLLDDN 120

OY 138 NTRGLDILLALDSHPNIEVLEFNPVLKRRALGYLTDFRLNRHMKSFADNRATI 197

DB 121 NTRGLDILLALDSHPNIEVLEFNPVLKRRALGYLTDFRLNRHMKSFADNRATI 180

OY 198 LCGRNIGDEYFKVEDYFADLDILATGSVYGEVSHDFRYWASHAHNATRIIRSG 254

DB 181 LCGRNIGDEYFKVEDYFADLDILATGSVYGEVSHDFRYWASHAHNATRIIRSG 237

RESULT 3

AAV75751

ID AAV75751 standard; Protein; 507 AA.

XX AAV75751;

XX 21-MAR-2000 (first entry)

XX Neisseria gonorrhoeae ORF 987 protein sequence SEQ ID NO:2974.

DE Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.

XX Neisseria gonorrhoeae.

OS Neisseria gonorrhoeae.

XX WO957280-A2.

XX 11-NOV-1999.

PD 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR ) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;

DR WPI: 2000-062150/05.

DR N-PSDB: AAZ54513.

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT vaccines and diagnostics -

XX Claim 2: Page 1394; 1453pp; English.

PS AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAV74253 to AAV75941

CC represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides

CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

CC PCR primers used in the exemplification of the present invention. The

CC polypeptides, the polynucleotides, antibodies and compositions of

CC the invention can be used as vaccines, as diagnostic reagents, and as

CC immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to

CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves

CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.

SQ Sequence 507 AA: 41.0%; Score 215; DB 21; Length 507;  
 Query Match Best Local Similarity 99.7%; Pred. No. 1.9e-204;  
 Matches 315; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 157 VRLNPNVLRKRALGYLTDFPRLNRMHNSFTADNRATILGGRNIGDEYKVEDTVF 216  
 |||||  
 DB 139 VRLNPNVLRKRALGYLTDFPRLNRMHNSFTADNRATILGGRNIGDEYKVEDTVF 198  
 |||||

QY 217 ADDILTLGTSVGVSHDFPRYMSAHNATRIIRSGNIGKGLQALGYNDERTSHALLR 276  
 |||||  
 DB 199 ADDILTLGTSVGVSHDFPRYMSAHNATRIIRSGNIGKGLQALGYNDERTSHALLR 258  
 |||||

QY 277 YRETFEGSPLYOKIQTGRIDMQSVQTRILISDPKAGLDRRRKPPINGRLQDALKOPEKS 336  
 |||||  
 DB 259 YRETFEGSPLYOKIQTGRIDMQSVQTRILISDPKAGLDRRRKPPINGRLQDALKOPEKS 318  
 |||||

QY 337 VYLSPFVFPKSGTDALAKLVQDGIYVYLTNSLQATDVAAVHSGVYKRPKLKAGIK 396  
 |||||  
 DB 319 VYLSPFVFPKSGTDALAKLVQDGIYVYLTNSLQATDVAAVHSGVYKRPKLKAGIK 378  
 |||||

QY 397 LYELQPMHAPATKDKGLTSSVTSIAKTFIYDGKRIFGSFNLDRSATLNTMGVVI 456  
 |||||  
 DB 379 LYELQPMHAPATKDKGLTSSVTSIAKTFIYDGKRIFGSFNLDRSATLNTMGVVI 438  
 |||||

QY 457 ESPKIAEQMERTLADT 472  
 |||||  
 DB 439 ESPKIAEQMERTLADT 454  
 |||||

RESULT 4  
 ABP34270 ID ABP34270 standard; Protein; 71 AA.  
 AC ABP34270:  
 DT 08-JUL-2002 (first entry)  
 XX Human ORF3243 protein, SEQ ID NO:6486.  
 XX Human ORF3243 protein, SEQ ID NO:6486.  
 KW Human; ORF; open reading frame; ORF; drug screening; diagnosis;  
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;  
 KW immune modulation; haematopoiesis regulation; tissue growth;  
 KW angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;  
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;  
 KW behaviour; cancer; proliferative disorder; neurological disorder;  
 KW cardiovascular disease; immune system disorder; organ transplantation;  
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnerability;  
 KW vasotropic; antiproliferative; antidiabetic; cytostatic; neurotropic;  
 KW neuroprotective; antithrombotic; antithrombotic; thrombolytic;  
 KW cardiant; hypotensive; antihypertensive; antihypertensive; immunomodulator;  
 KW dermatological; analgesic; virucide; antibacterial; fungicide.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 PN WO200190366-A2.  
 PD 29-NOV-2001.  
 PF 24-MAY-2001; 2001MO-US1076.  
 XX 24-MAY-2001; 2001MO-US1076.  
 PR 24-MAY-2000; 2000US-206690P.  
 XX (CURA-) CURAGEN CORP.  
 PA Leach MD, Shinkets RA.  
 PI Leach MD, Shinkets RA.  
 DR WPI; 2002-106200/14.  
 DR N-PSDB; ABN78236.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and disorders related to organ  
 PT transplantation  
 XX Claim 10; Page 1869; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
 CC ABN75058 represent cDNAs encoding them. The invention also encompasses  
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
 CC polypeptides, methods of screening for modulators of ORFX expression or  
 CC activity, and methods of screening individuals for a predisposition to an  
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
 CC range of biological activities, such as cytokine, cell proliferation,  
 CC cell differentiation, immune modulation, haematopoiesis regulation,  
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
 CC and antiinfective activity, and may also be involved in the determination  
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
 CC nucleic acids and antibodies may be used in the treatment of cancers,  
 CC other proliferative disorders such as psoriasis and benign tumours,  
 CC neurological disorders such as epilepsy and Alzheimer's disease,  
 CC cardiovascular diseases, immune system disorders, disorders related to  
 CC organ transplantation, disorders of tissue growth and regeneration,  
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
 CC storage disease, and infectious diseases caused by viral, bacterial,  
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
 CC source of primers and probes. In the detection of ORFX genomic sequences  
 CC or transcripts, in the identification and cloning of homologous  
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
 CC nucleic acids may additionally be used to produce transgenic animals  
 CC which may be useful for studying the function and/or activity of ORFX  
 CC protein, and in drug screening. The ORFX proteins may also be used as  
 CC immunogens to generate specific antibodies, which are useful in the  
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
 CC XX

SQ Sequence 71 AA;  
 Query Match Best Local Similarity 100.0%; Score 8; DB 23; Length 71;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 ALGYLTDF 177  
 |||||  
 DB 47 ALGYLTDF 54  
 |||||

RESULT 5  
 AAU62609 ID AAU62609 standard; Protein; 100 AA.  
 AC AAU62609:  
 DT 27-FEB-2002 (first entry)  
 XX Propionibacterium acnes immunogenic protein #23505.  
 DE Propionibacterium acnes immunogenic protein #23505.  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX Propionibacterium acnes.  
 OS Propionibacterium acnes.  
 XX WO200181581-A2.  
 PN 01-NOV-2001.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,



PF 20-APR-2001; 2001MO-US12865.  
 XX 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Skelley YAM, Persing DH, Mitcham JL, Wang SS, Bhalla A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX WPI; 2001-616774/71.  
 DR N-PSDB; AAS59627.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 XX Example 1; SEQ ID NO 23804; 10699p; English.  
 PS  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence. For example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SO Sequence 100 AA;  
 Query Match 1.5%; Score 8; DB 22; Length 100;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 198 LGGRNIGD 205  
 Db 33 LGGRNIGD 40  
 RESULT 6  
 ID AAG91101 standard; Protein: 498 AA.  
 XX  
 AC AAG91101;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 4855.  
 XX  
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KM organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX

PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW) KYOWA HAKKO KOGYO KK.  
 PA  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH6320.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 17; SEQ ID NO: 4855; 246pp + Sequence listing; English.  
 CC  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 CC  
 SO Sequence 498 AA;  
 Query Match 1.5%; Score 8; DB 22; Length 498;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 440 NLDPARSAR 447  
 Db 418 NLDPARSAR 425  
 RESULT 7  
 ID AAU15087 standard; Protein: 579 AA.  
 XX  
 AC AAU15087;  
 XX  
 DT 04-DEC-2001 (first entry)  
 XX  
 DE Protein encoded by C. albicans essential gene CATLR103C (CDC45).  
 XX  
 KW Gene identification; essential gene; GRACE; pathogenic fungus;  
 KM gene replacement and conditional expression; fungal infection.  
 XX  
 OS Candida albicans.  
 XX  
 PN W0200160975-A2.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 20-FEB-2001; 2001WO-US05551.  
 XX  
 PR 18-FEB-2000; 2000US-0183534.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Roemer T, Jiang B, Boone C, Bussey H;  
 XX WPI; 2001-489080/53.  
 DR N-PSDB; AAS23415.  
 DR

```

XX Identifying genes essential to fungal metabolisms and identifying
PT potential therapeutic agents that target these genes
XX
PS Claim 43: Page 217-218; 324pp; English.
XX
CC The present invention relates to novel methods for constructing fungal
CC strains useful for identification and validation of gene products as
CC targets for therapeutic agents, for creating a collection of identified
CC essential genes, and screening assays for the discovery of new drugs.
CC The invention provides the GRACE (gene replacement and conditional
CC expression) method for the construction of mutant organisms referred to
CC as GRACE strains of the organism. The invention can be applied to any
CC organism, particularly a pathogenic fungus e.g. Candida albicans,
CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
CC useful to identify agents that may be used in the treatment of fungal
CC infections. AAU15053-AAU15113 represent proteins encoded by C. albicans
CC essential genes.
XX
SQ Sequence 579 AA;
XX
Query Match 1.5%; Score 8; DB 22; Length 579;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 512 AAKITSL 519
DB 41 AAKITSL 48
XX
RESULT 8
AAB79422
ID AAB79422 standard; Protein: 564 AA.
XX
AC AAB79422:
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:360.
XX
KW Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diptheriae; evolutionary study.
XX
XX
OS Corynebacterium glutamicum.
XX
XX
XX WO200100844-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000MO-IB00943.
XX
XX 25-JUN-1999; 99US-0141031.
XX 08-JUL-1999; 99DE-1031412.
XX 08-JUL-1999; 99DE-1031413.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031431.
XX 08-JUL-1999; 99DE-1031433.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031562.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1032180.
XX 08-JUL-1999; 99DE-1032227.
XX 08-JUL-1999; 99DE-1032230.
XX 09-JUL-1999; 99US-0143208.

```

```

PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Habernauer G;
PI WPT: 2001-061975/07.
XX
XX N-PSDB: AAF71539.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes.
XX
XX Claim 20; Page 646-648; 1246pp; English.
XX
PS AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
XX
XX metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX
XX
SQ Sequence 584 AA;
XX
XX
Query Match 1.5%; Score 8; DB 22; Length 584;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 440 NLDPDSAR 447
DB 504 NLDPDSAR 511
XX
XX
RESULT 9
ABG06295
ID ABG06295 standard; Protein: 683 AA.
XX
XX
XX ABG06295;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #6286.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX

```

PN WO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB; AAS70482.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID No 36654; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ Sequence 683 AA;  
Query Match 1.5%; Score 8; DB 22; Length 683;  
Best Local Similarity 100.0%; Pred. No. 81;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 TDVAAVHS 381  
DB 437 TDVAAVHS 444  
RESULT 10  
ABG08784  
ID ABG08784 standard; Protein: 692 AA.  
XX  
XX ABG08784;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #8775.  
DE  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX

PD 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB; AAS72971.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 20; SEQ ID No 39143; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
SQ Sequence 692 AA;  
Query Match 1.5%; Score 8; DB 22; Length 692;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 374 TDVAAVHS 381  
DB 446 TDVAAVHS 453  
RESULT 11  
ABG01448  
ID ABG01448 standard; Protein: 796 AA.  
XX  
XX ABG01448;  
AC  
XX  
XX 13-FEB-2002 (first entry)  
DT  
XX  
XX Novel human diagnostic protein #1439.  
DE  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200175067-A2.  
PN  
XX  
XX 11-OCT-2001.  
PD  
XX

PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HXSEQ INC.
PI	Drimaeac RT, Liu C, Tang YT;
DR	WPI: 2001-639362/73.
DR	N-PSTDB; AAS75984.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID No 42156; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pcl_sequences.
XX	
SQ	Sequence 796 AA:
OY	Query Match 1.5%; Score 8; DB 22; Length 796; Best Local Similarity 100.0%; Pred. No. 94;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Db	374 TDVAAVHS 381         437 TDVAAVHS 444
RESULT 13	
ABG05821	ID ABG05821 standard; Protein: 882 AA.
XX	ABG05821;
AC	
DT	13-FEB-2002 (first entry)
DE	Novel human diagnostic protein #5812.
XX	
KW	Hunan; chromosome mapping; gene mapping; forensic;
XX	food supplement; medical imaging; diagnostic; genetic disorder.
OS	Homo sapiens.
PN	WO200175067-A2.
PD	
PF	11-OCT-2001.
PP	
PR	30-MAR-2001; 2001MO-US08631.
PR	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PR	

```

XX (HYSE-) HYSEQ INC.
PA
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAO13288.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 20: SEQ ID No 36180; 103pp: English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 882 AA:
SO
Query Match 1.5%; Score 8; DB 22; Length 882;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3/4 TDVAAVHS 381
DB 437 TDVAAVHS 444

```

```

PI Thibaut D;
XX
XX WPI: 1991-252650/34.
DR N-PSDB; AAO13288.
XX
XX New polypeptide(s) involved in cobalamin and cobamide
PT biosynthesis - and DNA encoding them, for amplification of
PT cobalamin, esp. coenzyme B12 prodn.
XX
XX Claim 15; Fig 47; 299pp: French.
PS
XX This sequence corresponds to one of 24 polypeptides obtained from
CC P.dentrificans and implicated in the biosynthesis of cobalamines
CC and/or cobamides. It is encoded by part of a 13144 bp fragment
CC isolated from a P.dentrificans genomic DNA bank constructed in
CC vector pXL59.
XX See also AAO13284-Q13287.
XX
XX Sequence 1275 AA:
SO
Query Match 1.5%; Score 8; DB 12; Length 1275;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 388 KPLKAGI 395
DB 170 KPLKAGI 177

```

```

RESULT 15
AAR49463
ID AAR49463 standard; Protein; 18 AA.
XX
XX AAR49463;
AC
XX
XX 16-SEP-1994 (first entry)
DT
XX
XX HLA-DR alpha-chain position 111-128.
DE
XX
XX Naturally-occurring; immunomodulatory protein; human; therapy; class I;
KW major histocompatibility complex; class II; allotype; type I diabetes;
KW autoimmune disease; rheumatoid arthritis; T-cell-mediated response;
KW multiple sclerosis; transplant rejection; vaccine; MHC.
XX
XX Homo sapiens.
OS
XX
XX WO9404171-A.
PN
XX
XX 03-MAR-1994.
PD
XX
XX 11-AUG-1993; 93WO-US07545.
PF
XX
XX 11-AUG-1992; 92US-0925460.
PR
XX 15-JUN-1993; 93US-0925460.
XX
XX (HARD ) HARVARD COLLEGE.
PA
XX
XX Chicx RM, Hedley ML, Stern LJ, Strominger JL, Urban RG;
PI Vignali DA.
XX
XX WPI: 1994-082825/10.
DR
XX
XX Novel immunomodulatory peptide(s) and nucleic acids - useful for
PT treatment of autoimmune diseases, transplant rejection and for
PT vaccination
XX
XX Disclosure: Page 42; 139pp: English.
XX
XX The sequences given in AAR49291-505 and AAR46981-7038 represent peptide
CC fragments of naturally-occurring immunomodulatory proteins. These
CC fragments are between 10-30 residues in length and bind to a human
CC major histocompatibility complex (MHC) class II allotype. These
CC peptides may be used for therapy of autoimmune diseases, such as

```

CC type I diabetes, rheumatoid arthritis and multiple sclerosis, and to  
 CC reduce transplant rejection. They may also be used for vaccination  
 CC providing an exclusively T-cell-mediated response, which can be  
 CC class I or class-II based, or both, depending on the length and  
 CC character of the immunogenic peptides.

XX Sequence 18 AA:

Query Match 1.3%; Score 7; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 364 VTUVTNS 370  
 Db 4 VTUVTNS 10

RESULT 16

AAR49462 ID AAR49462 standard; Protein; 19 AA.

XX AAR49462;

DT 16-SEP-1994 (first entry)

DE HLA-DR alpha-chain position 111-129.

XX Naturally-occurring; immunomodulatory protein; human; therapy; class I;  
 KW major histocompatibility complex; class II; allotype; type I diabetes;  
 KW autoimmune disease; rheumatoid arthritis; T-cell-mediated response;  
 KW multiple sclerosis; transplant rejection; vaccine; MHC.

XX Homo sapiens.

OS WO9404171-A.

XX WO9404171-A.

XX 03-MAR-1994.

XX 11-AUG-1993; 93WO-US07545.

XX 11-AUG-1992; 92US-0925460.

XX 15-JUN-1993; 93US-0925460.

XX (HARD ) HARVARD COLLEGE.

XX Chicz RM, Hedley ML, Stern LJ, Strominger JL, Urban JG;  
 PI Vignali DA;

XX WPI; 1994-082825/10.

XX Novel immunomodulatory peptide(s) and nucleic acids - useful for  
 PT treatment of auto-immune diseases, transplant rejection and for  
 PT vaccination

PS Disclosure; Page 42; 139pp; English.

XX The sequences given in AAR49291-505 and AAR46981-7038 represent peptide  
 CC fragments of naturally-occurring immunomodulatory proteins. These  
 CC fragments are between 10-30 residues in length and bind to a human  
 CC major histocompatibility complex (MHC) class II allotype. These  
 CC peptides may be used for therapy of autoimmune diseases, such as  
 CC type I diabetes, rheumatoid arthritis and multiple sclerosis, and to  
 CC reduce transplant rejection. They may also be used for vaccination  
 CC providing an exclusively T-cell-mediated response, which can be  
 CC class I or class-II based, or both, depending on the length and  
 CC character of the immunogenic peptides.

XX Sequence 19 AA:

Query Match 1.3%; Score 7; DB 15; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 364 VTUVTNS 370  
 Db 4 VTUVTNS 10

RESULT 17

AAR26283 ID AAR26283 standard; peptide; 21 AA.

XX AAR26283;

DT 28-JAN-1993 (first entry)

DE Endothelin-2.

XX Analogue; endothelin-receptor; hypertension; cardiac;  
 KW cerebral; circulatory; asthma; vasodilator; ET.

XX Synthetic.

XX Key Location/Qualifiers

FT Disulfide-bond 1..15

FT Disulfide-bond 3..11

XX EP499266-A.

XX 19-AUG-1992.

XX 14-FEB-1992; 92EP-0102498.

XX 15-FEB-1991; 91JP-0022220.

XX 19-DEC-1991; 91JP-0337235.

XX (TAKE ) TAKEDA CHEM IND LTD.

XX Kikuchi T, Kubo K, Wakimasu M;

XX WPI; 1992-278006/34.

XX New endothelin antagonist peptide(s) for treating hypertension -  
 PT also useful as vasodilators and for cardiac, cerebral,  
 PT circulatory and renal diseases and asthma

XX Disclosure; Page 10; 39pp; English.

XX The peptide represents endothelin-2 (ET-2). Analogues of this  
 CC peptide may be produced which can bind to endothelin receptors  
 CC of warm-blooded animals but can not exhibit endothelin-like  
 CC constrictor activity. Thus they function as endothelin antagonists  
 CC and can be used as prophylactic and therapeutic drugs for hypertension,  
 CC cardiac or cerebral circulatory diseases (e.g. acute renal  
 CC insufficiency) and asthma. They may also be used as agents for  
 CC improving circulatory functions and as vasodilators and have  
 CC bronchial smooth muscle constrictor activity..  
 CC See also AAR26279-302.

XX Sequence 21 AA:

Query Match 1.3%; Score 7; DB 13; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7

RESULT 18

AAR30557 ID AAR30557 standard; peptide; 21 AA.

XX AAR30557;

DT	28-JAN-1993	(first entry)				
XX	Thr18, Leu19	Endothelin-2.				
DE						
KW	Analogue; endothelin-receptor; hypertension; cardiac;					
KM	cerebral; circulatory; asthma; vasodilator; ET.					
XX						
OS	Synthetic.					
XX						
PH	Key	Location/Qualifiers				
FT	Disulfide-bond	1..15				
FT	Disulfide-bond	3..11				
XX						
PN	EP49266-A.					
XX						
PD	19-AUG-1992.					
XX						
FE	14-FEB-1992;	92EP-0102498.				
XX						
PR	15-FEB-1991;	91JP-0022220.				
PR	19-DEC-1991;	91JP-0337235.				
XX						
PA	(TAKE ) TAKEKA CHEM IND LTD.					
XX						
PI	Kikuchi T, Kubo K, Wakimasu M;					
XX						
DR	WPI: 1992-278006/34.					
XX						
PT	New endothelin antagonist peptide(s) for treating hypertension -					
PT	also useful as vasodilators and for cardiac, cerebral,					
PT	circulatory and renal diseases and asthma					
XX						
PS	Claim 11; Page 11; 39pp: English.					
XX						
CC	The peptide represents an analogue of endothelin-2 (ET-2) having					
CC	the substns. Thr18 and Leu19. Such a peptide can bind to					
CC	endothelin receptors of warm-blooded animals but can not exhibit					
CC	endothelin- like constrictor activity. Thus they function as					
CC	therapeutic drugs for hypertension, cardiac or cerebral					
CC	circulatory diseases (e.g. acute renalinsufficiency) and asthma.					
CC	They may also be used as agents for improving circulatory functions					
CC	and as vasodilators and have bronchial smooth muscle constrictor					
CC	activity. See also AAR26279-302.					
XX						
XX	Sequence	21 AA:				
XX						
SO	Query Match	1.3%; Score 7; DB 13; Length 21;				
	Best Local Similarity	100.0%; Pred. No. 30;				
	Matches	7; Conservative	0; Mismatches	0; Indels	0; Gaps	0;
OY	32 CSCSSWL 38					
	1 CSCSSWL 7					
Db						
RESULT 19						
AAR38383						
ID	AAR38383 standard; peptide: 21 AA.					
XX						
AC	AAR38383;					
XX						
DT	27-OCT-1993 (first entry)					
XX						
DE	Endothelin receptor antagonist [Thr18,Leu19]-ET-2.					
XX						
KW	Treatment; cardiac infarction; acute renal insufficiency; asthma.					
XX						
OS	Synthetic.					
XX						
XX	Key	Location/Qualifiers				
PH	Disulfide-bond	1..15				
FT	Disulfide-bond	3..11				

[illegible]

XX Claim 5; Page 30; 40pp; German.  
 PS  
 CC This invention describes novel complexes of compounds which have the  
 CC formula E-L-(K)b with metal ions of atomic number 21-32, 37-39, 42-51  
 CC and 57-83 where E = residue of an endothelin (or its derivative, partial  
 CC sequence, analogue or antagonist), including forms with free SH groups  
 CC that bind directly to metal, optionally labelled with radioactive iodine  
 CC isotopes, L = bond, 21-R-22 or a group (1), R = 1-200, optionally  
 CC unsaturated alkyl, optionally substituted by OH and/or epoxy, and  
 CC optionally interrupted by 1 or more O, S, CO, NCO or NH (the last 2  
 CC isotopes, L = bond, 21-R-22 or a group (1), R = 1-200, optionally  
 CC NH, CO, CSO, NH, NHCOS, s and t = 0-3, B = phenyl or cyclohexyl  
 CC ring, b = 0 or 1, K is a chelating residue. The compounds of the  
 CC invention are useful as diagnostic agents, including contrast agents for  
 CC X-ray and NMR imaging of pathological vascular changes e.g.  
 CC atherosclerosis. They can also be used as radiopharmaceuticals. The  
 CC invention allows the marking of possible early non-invasive diagnosis  
 CC of vascular changes before these become stenotic. The E component  
 CC provides strong localisation to arterial wall lesions without significant  
 CC accumulation in other tissues or organs. Compared with known diagnostic  
 CC agents they have better pharmacokinetic properties and provide improved  
 CC contrast. They have excellent in vitro and in vivo stability (no  
 CC significant release or exchange of bound radionuclides). This sequence  
 CC is a specific claimed example of a peptide used in the construction of  
 CC the metalcomplex-conjugates of the invention.  
 CC  
 XX Sequence 21 AA:  
 SO  
 Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7  
 RESULT 21  
 AAY08514  
 ID AAY08514 standard; peptide: 21 AA.  
 XX  
 AC AAY08514;  
 XX  
 DT 30-JUL-1999 (first entry)  
 XX  
 DE Endothelin metal conjugate peptide 13.  
 XX  
 KW Endothelin; metal conjugate; complex; diagnostic; contrast agent;  
 KW X-ray imaging; NMR imaging; pathological vascular change; stenotic;  
 KW atherosclerosis; radiopharmaceutical; non-invasive diagnosis;  
 KW arterial wall lesion; pharmacokinetic.  
 XX  
 OS Synthetic.  
 XX  
 PN DE4301871-A1.  
 XX  
 PD 14-JUL-1994.  
 XX  
 PF 13-JAN-1993; 93DE-4301871.  
 XX  
 PR 13-JAN-1993; 93DE-4301871.  
 XX  
 PA (DIAG-) INST DIAGNOSTIKFORSCHUNG GMBH.  
 PA (DINK/) DINKELBORG L.  
 PA (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 XX  
 PI Dinkelborg L, Erber S, Gries H, Hilger CS, Kramp W;  
 PI Platzek J, Schlier H, Speck U, Reiser JH;  
 DR WPI; 1994-226321/28.  
 XX  
 PT New complexes of derivatised endothelin or related cpds. - with

PT radioactive or paramagnetic metal ions, useful as diagnostic agents  
 PT for imaging pathological changes in blood vessels  
 XX  
 PS Claim 5; Page 30; 40pp; German.  
 CC This invention describes novel complexes of compounds which have the  
 CC formula E-L-(K)b with metal ions of atomic number 21-32, 37-39, 42-51  
 CC and 57-83 where E = residue of an endothelin (or its derivative, partial  
 CC sequence, analogue or antagonist), including forms with free SH groups  
 CC that bind directly to metal, optionally labelled with radioactive iodine  
 CC isotopes, L = bond, 21-R-22 or a group (1), R = 1-200, optionally  
 CC unsaturated alkyl, optionally substituted by OH and/or epoxy, and  
 CC optionally interrupted by 1 or more O, S, CO, NCO or NH (the last 2  
 CC isotopes, L = bond, 21-R-22 or a group (1), R = 1-200, optionally  
 CC NH, CO, CSO, NH, NHCOS, s and t = 0-3, B = phenyl or cyclohexyl  
 CC ring, b = 0 or 1, K is a chelating residue. The compounds of the  
 CC invention are useful as diagnostic agents, including contrast agents for  
 CC X-ray and NMR imaging of pathological vascular changes e.g.  
 CC atherosclerosis. They can also be used as radiopharmaceuticals. The  
 CC invention allows the marking of possible early non-invasive diagnosis  
 CC of vascular changes before these become stenotic. The E component  
 CC provides strong localisation to arterial wall lesions without significant  
 CC accumulation in other tissues or organs. Compared with known diagnostic  
 CC agents they have better pharmacokinetic properties and provide improved  
 CC contrast. They have excellent in vitro and in vivo stability (no  
 CC significant release or exchange of bound radionuclides). This sequence  
 CC is a specific claimed example of a peptide used in the construction of  
 CC the metalcomplex-conjugates of the invention.  
 CC  
 XX Sequence 21 AA:  
 SO  
 Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7  
 RESULT 22  
 AAR60077  
 ID AAR60077 standard; Protein: 21 AA.  
 XX  
 AC AAR60077;  
 XX  
 DT 21-MAR-1995 (first entry)  
 XX  
 DE Endothelin-2.  
 XX  
 KW Endothelin; ET; ET-1; ET-2; ET-3; pentapeptide; receptor;  
 KW antagonist; myocardial infarction; hypertension;  
 KW acute renal insufficiency; asthma.  
 XX  
 OS Synthetic.  
 XX  
 PN JP06192293-A.  
 XX  
 PD 12-JUL-1994.  
 XX  
 PF 20-JUN-1991; 91JP-0148807.  
 XX  
 PR 20-JUN-1991; 91JP-0148807.  
 XX  
 PA (TAKE ) TAKEDA CHEM. IND LTD.  
 PA  
 DR WPI; 1994-260512/32.  
 XX  
 PT Cyclic penta:peptide(s) - useful as endothelin receptor



PT antagonists for treatment of myocardial infarction, etc.  
 XX  
 PS Disclosure; Page 2; 13pp; Japanese.  
 XX  
 CC The sequences of ET-1, ET-2 and ET-3 are given in AAR60076-78. Cyclic  
 CC pentapeptides, useful as endothelin receptor antagonists are given  
 CC in AAR60079-88. They are useful in the treatment of prophylaxis of  
 CC hypertension, myocardial infarction, acute renal insufficiency or  
 CC asthma.  
 XX  
 SO Sequence 21 AA;  
 Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7  
 RESULT 23  
 AAR62458  
 ID AAR62458 standard; peptide: 21 AA.  
 AC AAR62458;  
 XX  
 DT 19-MAY-1995 (first entry)  
 XX  
 DE Endothelin peptide conjugated to chalcogen-contg. ligand.  
 XX  
 KM Chalcogen; ligand; complex; endothelin; non-invasive imaging;  
 KM technetium; rhenium; in vivo imaging.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 21  
 FT /note= "attached to chalcogen-contg. ligand,  
 FT see comments"  
 XX  
 PN DE4311022-A.  
 XX  
 PD 06-OCT-1994.  
 XX  
 PF 31-MAR-1993; 93DE-4311022.  
 XX  
 PR 31-MAR-1993; 93DE-4311022.  
 XX  
 PA (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 XX  
 PI Dinkelborg L, Hilger C, Kramp W, Schlier H;  
 XX  
 DR WPI; 1994-311304/39.  
 XX  
 XX Bis-(N-(mercapto-ethyl- or -propyl) amino:carbonyl:methyl) oxide,  
 PT sulphide or selenide and derivs. - are ligands for radioactive  
 PT technetium or rhenium for diagnosis and therapy, e.g. for  
 PT non-invasive in-vivo visualisation of receptors  
 XX  
 PS Claim 9; Page 13; 15pp; German.  
 XX  
 CC This peptide is one of 14 endothelin derivatives which are  
 CC preferred for use in conjugates with novel bis-(N-(mercapto-ethyl-  
 CC or -propyl) aminocarbonylmethyl) oxide, sulphide or selenide  
 CC compounds. The chalcogen-containing compounds are ligands which are  
 CC complexed with a technetium or rhenium radioisotope. The  
 CC ligand-peptide conjugates are useful for non-invasive, in vivo  
 CC imaging, e.g. of receptors.  
 CC  
 SO Sequence 21 AA;  
 Query Match 1.3%; Score 7; DB 15; Length 21;

Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7  
 RESULT 24  
 AAR62451  
 ID AAR62451 standard; peptide: 21 AA.  
 AC AAR62451;  
 XX  
 DT 19-MAY-1995 (first entry)  
 XX  
 DE Endothelin peptide conjugated to chalcogen-contg. ligand.  
 XX  
 KM Chalcogen; ligand; complex; endothelin; non-invasive imaging;  
 KM technetium; rhenium; in vivo imaging.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 1..15  
 FT Disulfide-bond 3..11  
 FT Modified-site 21  
 FT /note= "attached to chalcogen-contg. ligand,  
 FT see comments"  
 XX  
 PN DE4311022-A.  
 XX  
 PD 06-OCT-1994.  
 XX  
 PF 31-MAR-1993; 93DE-4311022.  
 XX  
 PR 31-MAR-1993; 93DE-4311022.  
 XX  
 PA (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 XX  
 PI Dinkelborg L, Hilger C, Kramp W, Schlier H;  
 XX  
 DR WPI; 1994-311304/39.  
 XX  
 XX Bis-(N-(mercapto-ethyl- or -propyl) amino:carbonyl:methyl) oxide,  
 PT sulphide or selenide and derivs. - are ligands for radioactive  
 PT technetium or rhenium for diagnosis and therapy, e.g. for  
 PT non-invasive in-vivo visualisation of receptors  
 XX  
 PS Claim 9; Page 13; 15pp; German.  
 XX  
 CC This peptide is one of 14 endothelin derivatives which are  
 CC preferred for use in conjugates with novel bis-(N-(mercapto-ethyl-  
 CC or -propyl) aminocarbonylmethyl) oxide, sulphide or selenide  
 CC compounds. The chalcogen-containing compounds are ligands which are  
 CC complexed with a technetium or rhenium radioisotope. The  
 CC ligand-peptide conjugates are useful for non-invasive, in vivo  
 CC imaging, e.g. of receptors.  
 CC  
 SO Sequence 21 AA;  
 Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 CSCSSWL 38  
 Db 1 CSCSSWL 7  
 RESULT 25  
 AAR65730  
 ID AAR65730 standard; peptide: 21 AA.

```

XX AC AAR65730;
XX XX
XX DT 02-JUN-1995 (first entry)
XX XX
DE Endothelin peptide conjugated to chalcogen-contg. bifunctional ligand.
XX XX
XX KW Chalcogen; ligand; complex; endothelin; non-invasive imaging;
XX KW technetium; rhenium; in vivo imaging; atherosclerotic plaque;
XX KW bifunctional chalcogen-interrupted chelate former.
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT Disulfide-bond 1..15
XX FT Disulfide-bond 3..11
XX FT Modified-site 21
XX FT /note= "attached to chalcogen-contg. ligand,
XX FT see comments"
XX FT
XX FT DE4311023-A.
XX PN
XX PD 06-OCT-1994.
XX XX
XX PF 31-MAR-1993; 93DE-4311023.
XX PR
XX PR 31-MAR-1993; 93DE-4311023.
XX PA
XX PI (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.
XX PI Dinkelborg L, Hilger C, Kramp W, Schlier H;
XX DR WPI; 1994-311305/39.
XX XX
XX PT Bifunctional chalcogen-atom interrupted chelate formers - for
XX PT radioactive isotopes, their prodn. and use in diagnostics and
XX PT therapy, e.g. for non-invasive in-vivo representation of
XX PT receptors
XX PT
XX PS Claim 9; Page 11; 13pp; German.
XX PS
XX CC This peptide is one of 14 endothelin derivatives which are
XX CC preferred for use in conjugates with novel, bifunctional chalcogen-
XX CC interrupted chelate forming compounds. The chalcogen-containing
XX CC compounds are ligands which are complexed with a technetium or
XX CC rhenium radioisotope. The ligand-peptide conjugates are useful
XX CC for non-invasive, in vivo imaging, e.g. of receptors or of
XX CC atherosclerotic plaques.
XX CC
XX SO Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38
DB 1 CSCSSWL 7

RESULT 26
AAR65737
ID AAR65737 standard; peptide; 21 AA.
XX
XX AAR65737;
XX
XX DT 09-JUN-1995 (first entry)
XX XX
XX DE Endothelin peptide conjugated to chalcogen-contg. bifunctional ligand.
XX XX
XX KW Chalcogen; ligand; complex; endothelin; non-invasive imaging;
XX KW technetium; rhenium; in vivo imaging; atherosclerotic plaque;
XX KW bifunctional chalcogen-interrupted chelate former.

```

```

XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT Modified-site 21
XX FT /note= "attached to chalcogen-contg. ligand,
XX FT see comments"
XX FT
XX FT DE4311023-A.
XX PN
XX PD 06-OCT-1994.
XX XX
XX PF 31-MAR-1993; 93DE-4311023.
XX PR
XX PR 31-MAR-1993; 93DE-4311023.
XX PA
XX PI (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.
XX PI Dinkelborg L, Hilger C, Kramp W, Schlier H;
XX DR WPI; 1994-311305/39.
XX XX
XX PT Bifunctional chalcogen-atom interrupted chelate formers - for
XX PT radioactive isotopes, their prodn. and use in diagnostics and
XX PT therapy, e.g. for non-invasive in-vivo representation of
XX PT receptors
XX PT
XX PS Claim 9; Page 11; 13pp; German.
XX PS
XX CC This peptide is one of 14 endothelin derivatives which are
XX CC preferred for use in conjugates with novel, bifunctional chalcogen-
XX CC interrupted chelate forming compounds. The chalcogen-containing
XX CC compounds are ligands which are complexed with a technetium or
XX CC rhenium radioisotope. The ligand-peptide conjugates are useful
XX CC for non-invasive, in vivo imaging, e.g. of receptors or of
XX CC atherosclerotic plaques.
XX CC
XX SO Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38
DB 1 CSCSSWL 7

RESULT 27
AAR65716
ID AAR65716 standard; peptide; 21 AA.
XX
XX AAR65716;
XX
XX DT 02-JUN-1995 (first entry)
XX XX
XX DE Endothelin peptide conjugated to chalcogen-contg. bifunctional ligand.
XX XX
XX KW Chalcogen; ligand; complex; endothelin; non-invasive imaging;
XX KW technetium; rhenium; in vivo imaging; atherosclerotic plaque;
XX KW bifunctional chalcogen-interrupted chelate former.
XX KW
XX OS Synthetic.
XX XX
XX FH Key Location/Qualifiers
XX FT Disulfide-bond 1..15
XX FT Disulfide-bond 3..11
XX FT Modified-site 21
XX FT /note= "attached to chalcogen-contg. ligand,
XX FT see comments"
XX FT
XX FT DE4310999-A.
XX PN
XX

```

PD 06-OCT-1994.  
 XX 31-MAR-1993; 93DE-4310999.  
 XX 31-MAR-1993; 93DE-4310999.  
 XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 PA Dinkelborg L, Hilger C, Kramp W, Schlier H;  
 XX WPI; 1994-311297/39.  
 DR  
 XX Bifunctional, chalcogen-interrupted chelate former for  
 PT radioisotopes - used in diagnostics and therapy, e.g. for  
 PT non-invasive in-vivo representation of receptors  
 PS Claim 9; Page 11; 13pp; German.  
 XX  
 CC This peptide is one of 14 endothelin derivatives which are  
 CC preferred for use in conjugates with novel, bifunctional chalcogen-  
 CC interrupted chelate forming compounds. The chalcogen-containing  
 CC compounds are ligands which are complexed with a technetium or  
 CC rhodium radioisotope. The ligand-peptide conjugates are useful  
 CC for non-invasive, in vivo imaging, e.g. of receptors or of  
 CC atherosclerotic plaques.  
 CC  
 SQ Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 Db 1 CSCSSWL 7

RESULT 28  
 AAR65723  
 ID AAR65723 standard; peptide; 21 AA.

AC AAR65723;

DT 02-JUN-1995 (first entry)

XX Endothelin peptide conjugated to chalcogen-contg. bifunctional ligand.

KW Chalcogen; ligand; complex; endothelin; non-invasive imaging;  
 KW technetium; rhodium; in vivo imaging; atherosclerotic plaque;  
 KW bifunctional chalcogen-interrupted chelate former.  
 XX  
 OS Synthetic.

FT Key Location/Qualifiers  
 FT Modified-site 21 /note="attached to chalcogen-contg. ligand,  
 FT see comments"

PN DE4310999-A.

XX 06-OCT-1994.

XX 31-MAR-1993; 93DE-4310999.

XX 31-MAR-1993; 93DE-4310999.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

PI Dinkelborg L, Hilger C, Kramp W, Schlier H;

DR WPI; 1994-311297/39.

PT Bifunctional, chalcogen-interrupted chelate former for

PT radioisotopes - used in diagnostics and therapy, e.g. for  
 PT non-invasive in-vivo representation of receptors  
 XX  
 PS Claim 9; Page 11; 13pp; German.  
 XX  
 CC This peptide is one of 14 endothelin derivatives which are  
 CC preferred for use in conjugates with novel, bifunctional chalcogen-  
 CC interrupted chelate forming compounds. The chalcogen-containing  
 CC compounds are ligands which are complexed with a technetium or  
 CC rhodium radioisotope. The ligand-peptide conjugates are useful  
 CC for non-invasive, in vivo imaging, e.g. of receptors or of  
 CC atherosclerotic plaques.  
 CC  
 SQ Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 Db 1 CSCSSWL 7

RESULT 29  
 AAR63178  
 ID AAR63178 standard; peptide; 21 AA.

AC AAR63178;

DT 16-JUN-1995 (first entry)

XX Peptide ligand #9 for radiopharmaceutical agent.

KW Peptide; ligand; radioactive; peptide complex; imaging; receptor;  
 KW steroid; hormone; atherosclerotic plaque; targeting; endothelin;  
 KW analogue; derivative; antagonist.  
 XX  
 OS Synthetic.

PN WO9422491-A.

PD 13-OCT-1994.

PF 29-MAR-1994; 94WO-DE00369.

PR 31-MAR-1993; 93DE-4311021.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

PI Dinkelborg L, Erber S, Noll B, Rohlfes G, Schulze P;

DR WPI; 1994-332837/41.

XX New radio-pharmaceutical agents deriv. from peptide chelating  
 PT ligands - also the new ligands and complexes with tissue  
 PT targeting peptide(s), useful for imaging and treatment of e.g.  
 PT atherosclerosis or steroid dependent cancer  
 XX  
 PS Claim 14; Page 66; 81pp; German.

XX The sequences given in AAR73170-83 represent peptide ligands which are  
 CC used with the radioactive peptide complex of the invention. The  
 CC complexes may be used to image receptors, specifically of steroid  
 CC hormones, receptor contg. tissue and/or atherosclerotic plaque.  
 CC These peptides allow the complex to be targeted to a specific  
 CC tissue. These peptides are pref. derived from endothelin, and  
 CC represent analogues, derivatives or antagonists.  
 CC  
 SQ Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
 |||||  
 Db 1 CSCSSWL 7

RESULT 30  
 AAR63171  
 ID AAR63171 standard; peptide: 21 AA.  
 XX  
 AC AAR63171;

DF 16-JUN-1995 (first entry)

DE Peptide ligand #2 for radiopharmaceutical agent.

KW Peptide; ligand; radioactive; peptide complex; imaging; receptor;  
 KW steroid; hormone; atherosclerotic plaque; targeting; endothelin;  
 KW analogue; derivative; antagonist.

XX Synthetic.

XX Key location/Qualifiers

FT Disulfide-bond 1..15

XX Disulfide-bond 3..11

XX WO9422491-A.

XX 13-OCT-1994.

XX 29-MAR-1994; 94WO-DE00369.

XX 31-MAR-1993; 93DE-4311021.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

XX Dinkelborg L, Erber S, Noll B, Rohls G, Schulze P;  
 DR WPI; 1994-332837/41.

XX New radio-pharmaceutical agents deriv. from peptide chelating  
 PT ligands - also the new ligands and complexes with tissue  
 PT targeting peptide(s), useful for imaging and treatment of e.g.  
 PT atherosclerosis or steroid dependent cancer

XX Claim 14; Page 65; 81pp; German.

XX The sequences given in AAR63170-83 represent peptide ligands which are  
 CC used with the radioactive peptide complex of the invention. The  
 CC complexes may be used to image receptors, specifically of steroid  
 CC hormones, receptor contg. tissue and/or atherosclerotic plaque.  
 CC These peptides allow the complex to be targeted to a specific  
 CC tissue. These peptides are pref. derived from endothelin/ and  
 CC represent analogues, derivatives or antagonists.

XX Sequence 21 AA;

Query Match 1.3%; Score 7; DB 15; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
 |||||  
 Db 1 CSCSSWL 7

RESULT 31

ID AAR69409 standard; peptide: 21 AA.

XX AAR69409;  
 XX  
 XX

DF 16-JUL-1995 (first entry)  
 XX  
 DE [Thrl8, Leu19]-ET-2 for treating organ hypofunction.  
 XX  
 KW Endothelin receptor antagonist; organ hypofunction; liver; surgery;  
 KW transplant; cyclic.

XX Synthetic.

XX EP626174-A.

XX 30-NOV-1994.

XX 20-APR-1994; 94EP-0106093.

XX 21-APR-1993; 93JP-0094332.

XX (TAKE ) TAKEDA CHEM IND LTD.

XX Kitada C, Wakimasu M, Watanabe T;

XX WPI; 1995-000740/01.

PT Use of endothelin receptor antagonists - for prophylaxis or  
 PT treatment of organ hypofunction, partic. caused by surgery on or  
 PT transplant of the organ

XX Claim 9; Page 86; 107pp; English.

CC The invention relates to the new use of known endothelin receptor  
 CC antagonists for prophylaxis and/or therapy of hypofunction of organs  
 CC (particularly the liver) which occurs following their surgery or  
 CC transplant. The present sequence is a specific example of a suitable  
 CC antagonist listed in the Claims.

XX Sequence 21 AA;

Query Match 1.3%; Score 7; DB 16; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
 |||||  
 Db 1 CSCSSWL 7

RESULT 32

ID AAR86448 standard; peptide: 21 AA.

XX AAR86448;

XX 26-JAN-1996 (first entry)

XX Endothelin sequence for use as conjugate with Cys-free peptide.

XX diagnostic imaging; atherosclerotic plaque; tumour; inflammation;  
 KW conjugate; radiotherapy; cysteine-free; metal-binding; endothelin.

XX Synthetic.

XX DE4337599-A1.

XX 04-MAY-1995.

XX 01-NOV-1993; 93DE-4337599.

XX 01-NOV-1993; 93DE-4337599.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

XX Conrad J, Dinkelborg L, Erber S, Frommel C, Hohne W;  
 XX Kramp W, Kuttner G, Malin R, Schler HM, Schneider-Wergener J;  
 PI

PI Steindbrecher R, Froemmel C, Hoehne W, Kuetner G, Schier H;  
 XX WPI: 1995-171400/23.  
 DR  
 XX  
 PT New cysteine-free peptide(s) and their metal ion complexes - opt.  
 PT conjugated to targeting agent, useful for in vivo imaging of  
 PT tumours, atherosclerotic plaque etc. and for radiotherapy  
 PS  
 XX Claim 13; Page 20; 25pp; German.  
 CC New peptides of formula R1-X-R2 are provided, together with their  
 CC conjugates with peptides, proteins, biomolecules and macromolecules,  
 CC their complexes with metal ions and their water-soluble salts; where  
 CC X is a chain of up to 20 alpha, beta, and/or gamma amino acid residues,  
 CC including at least one Met, Arg, Lys or Asn but no Cys; R1 is H or a  
 CC hydrocarbyl or acyl group, or a bond to a conjugate; and R2 is OH,  
 CC hydrocarbyloxy or an optionally substituted amine group (forming an  
 CC amide), or a bond to a conjugate. The peptides are useful when complexed  
 CC to a metal (e.g. 99m-Tc) for in-vivo diagnostic imaging and therapy of  
 CC tumours, organs and foci of inflammation. Conjugates of the peptides can  
 CC be targeted to specific tissues. A particular application is imaging of  
 CC atherosclerotic plaque. Metal is bound with good in-vivo stability;  
 CC unbound material is rapidly cleared; and there is no allergic reaction.  
 CC Since the peptides are free of Cys, they can be prepared simply and  
 CC rapidly and can be presented in the form of a kit for radio-  
 CC pharmaceutical production. Specific examples of the peptides are  
 CC given in AAR86424 - AAR86439.  
 CC The present sequence is a specific example of an endothelin peptide  
 CC which can be used as a conjugate with the new peptides.  
 CC  
 XX Sequence 21 AA:  
 SQ  
 Query Match 1.3%; Score 7; DB 16; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 32 CSCSSWL 38  
 IIIIII  
 Db 1 CSCSSWL 7  
 RESULT 33  
 AAR86441  
 ID AAR86441 standard; peptide: 21 AA.  
 XX  
 AC AAR86441;  
 XX  
 XX 26-JAN-1996 (first entry)  
 DT  
 XX  
 DE Endothelin sequence for use as conjugate with Cys-free peptide.  
 XX  
 KW diagnostic imaging; atherosclerotic plaque; tumour; inflammation;  
 KW conjugate; radiotherapy; cysteine-free; metal-binding; endothelin.  
 XX  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Disulfide-bond 1..15  
 FT Disulfide-bond 3..11  
 XX  
 XX DE4337599-A1.  
 PN  
 XX  
 PD 04-MAY-1995.  
 XX  
 XX 01-NOV-1993; 93DE-4337599.  
 XX  
 XX 01-NOV-1993; 93DE-4337599.  
 PR  
 XX  
 XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 PA  
 XX Conrad J, Dinkelborg L, Erber S, Frommel C, Hoehne W,  
 PI Kramp W, Kuetner G, Malin R, Schier HM, Schneider-Meigener J,  
 PI Steindbrecher R, Froemmel C, Hoehne W, Kuetner G, Schier H;

XX  
 DR WPI: 1995-171400/23.  
 XX  
 PT New cysteine-free peptide(s) and their metal ion complexes - opt.  
 PT conjugated to targeting agent, useful for in vivo imaging of  
 PT tumours, atherosclerotic plaque etc. and for radiotherapy  
 PS  
 XX Claim 13; Page 19; 25pp; German.  
 CC New peptides of formula R1-X-R2 are provided, together with their  
 CC conjugates with peptides, proteins, biomolecules and macromolecules,  
 CC their complexes with metal ions and their water-soluble salts; where  
 CC X is a chain of up to 20 alpha, beta, and/or gamma amino acid residues,  
 CC including at least one Met, Arg, Lys or Asn but no Cys; R1 is H or a  
 CC hydrocarbyl or acyl group, or a bond to a conjugate; and R2 is OH,  
 CC hydrocarbyloxy or an optionally substituted amine group (forming an  
 CC amide), or a bond to a conjugate. The peptides are useful when complexed  
 CC to a metal (e.g. 99m-Tc) for in-vivo diagnostic imaging and therapy of  
 CC tumours, organs and foci of inflammation. Conjugates of the peptides can  
 CC be targeted to specific tissues. A particular application is imaging of  
 CC atherosclerotic plaque. Metal is bound with good in-vivo stability;  
 CC unbound material is rapidly cleared; and there is no allergic reaction.  
 CC Since the peptides are free of Cys, they can be prepared simply and  
 CC rapidly and can be presented in the form of a kit for radio-  
 CC pharmaceutical production. Specific examples of the peptides are  
 CC given in AAR86424 - AAR86439.  
 CC The present sequence is a specific example of an endothelin peptide  
 CC which can be used as a conjugate with the new peptides.  
 CC  
 XX Sequence 21 AA:  
 SQ  
 Query Match 1.3%; Score 7; DB 16; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 32 CSCSSWL 38  
 IIIIII  
 Db 1 CSCSSWL 7  
 RESULT 34  
 AAW33305  
 ID AAW33305 standard; peptide: 21 AA.  
 XX  
 AC AAW33305;  
 XX  
 XX 05-FEB-1998 (first entry)  
 DT  
 XX  
 DE Targeting conjugate for bis-nicotinamide bifunctional chelator.  
 XX  
 KW Bifunctional; bis-nicotinamide; chelator; radioactive; technetium;  
 KW rhodium; conjugate; endothelin; angiotensin; targeting agent;  
 KW organ; in vivo imaging; receptor; tissue; atherosclerotic plaque;  
 KW diagnosis; breast; prostatic; carcinoma; cytostatic agent;  
 KW delivery.  
 XX  
 XX Synthetic.  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FH Disulfide-bond 1..15  
 FT Disulfide-bond 3..11  
 XX  
 XX WO9710853-A2.  
 PN  
 XX  
 PD 27-MAR-1997.  
 XX  
 XX 19-SEP-1996; 96WO-DE01824.  
 PF  
 XX 21-SEP-1995; 95DE-4036783.  
 PR  
 XX  
 XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 PA  
 XX

PI Dinkelborg L, Hilger CS, Kramp W, Platzek J;  
 PI Raduechel B;  
 DR WPI; 1997-212562/19.  
 XX  
 PT New bis-nicotinamide bi:functional chelating agents - forming new,  
 PT stable technetium or rhenium complexes and conjugates useful as  
 PT radiodiagnostic and radio:therapeutic agents  
 XX  
 PS Claim 11; Page 40; 47pp; German.  
 XX

CC Novel bis-nicotinamide bifunctional chelating agent for radioactive  
 CC technetium and rhenium, when conjugated to an endothelin or  
 CC angiotensin derived targeting agent, e.g. the present peptide, can  
 CC be used for the in vivo imaging of organs, receptors and receptor  
 CC containing tissues and/or atherosclerotic plaques, e.g. for the  
 CC diagnosis of breast or prostatic carcinoma. They can also be used  
 CC for cytostatic agent delivery.  
 CC  
 SO Sequence 21 AA;

Query Match 1.3%; Score 7; DB 18; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 DB 1 CSCSSWL 7

RESULT 35  
 AAW32939  
 ID AAW32939 standard; peptide: 21 AA.  
 XX  
 AC AAW32939;

XX  
 DT 26-JAN-1998 (first entry)  
 XX

DE Targetting conjugate for bifunctional sulphonamide type ligand.

XX Bifunctional sulphonamide; ligand; radioactive; technetium;  
 KW rhenium; conjugate; endothelin; angiotensin; targeting agent;  
 KW organ; in vivo imaging; receptor; tissue; atherosclerotic plaque;  
 KW diagnosis; breast; prostatic; carcinoma; cytostatic agent;  
 KW delivery.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 1..15  
 FT Disulfide-bond 3..11  
 XX

XX DE19536785-A1.

XX PD 27-MAR-1997.

XX PF 21-SEP-1995; 95DE-1036785.

XX PR 21-SEP-1995; 95DE-1036785.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

XX PA Dinkelborg L, Erber S, Hilger CS, Kramp W, Platzek J;  
 PI Raduechel B;  
 PI

DR WPI; 1997-193894/18.

XX Bifunctional sulphonamide type ligands for radioactive technetium or  
 PT rhenium - and their chelates and conjugates, useful for imaging  
 PT carcinoma or atherosclerosis  
 XX  
 PS Claim 9; Page 11; 12pp; German.

XX  
 CC Novel bifunctional sulphonamide type ligands for radioactive  
 CC technetium and rhenium, when conjugated to an endothelin or  
 CC angiotensin derived targeting agent, e.g. the present peptide, can  
 CC be used for the in vivo imaging of organs, receptors and receptor  
 CC containing tissues and/or atherosclerotic plaques, e.g. for the  
 CC diagnosis of breast or prostatic carcinoma. They can also be used  
 CC for cytostatic agent delivery.  
 XX  
 SO Sequence 21 AA;

Query Match 1.3%; Score 7; DB 18; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 DB 1 CSCSSWL 7

RESULT 36  
 AAW11189  
 ID AAW11189 standard; peptide: 21 AA.  
 XX  
 AC AAW11189;

XX  
 DT 26-JAN-1998 (first entry)  
 XX

DE Targetting conjugate for bifunctional sulphonamide type ligand.

XX Bifunctional sulphonamide; ligand; radioactive; technetium;  
 KW rhenium; conjugate; endothelin; angiotensin; targeting agent;  
 KW organ; in vivo imaging; receptor; tissue; atherosclerotic plaque;  
 KW diagnosis; breast; prostatic; carcinoma; cytostatic agent;  
 KW delivery.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.

XX  
 FH Key Location/Qualifiers  
 FT Disulfide-bond 1..15  
 FT Disulfide-bond 3..11  
 XX

XX DE19536780-A1.

XX PD 27-MAR-1997.

XX PF 21-SEP-1995; 95DE-1036780.

XX PR 21-SEP-1995; 95DE-1036780.

XX (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.

XX PA Dinkelborg L, Erber S, Hilger CS, Kramp W, Platzek J;  
 PI Raduechel B;  
 PI

DR WPI; 1997-193892/18.

XX Bifunctional sulphonamide type ligands for radioactive technetium  
 PT and rhenium - and their chelates and conjugates with targeting  
 PT agents, useful for imaging carcinoma, atherosclerotic plaque, etc.  
 XX  
 PS Claim 11; Page 18; 19pp; German.  
 XX

CC Novel bifunctional sulphonamide type ligands for radioactive  
 CC technetium and rhenium, when conjugated to an endothelin or  
 CC angiotensin derived targeting agent, e.g. the present peptide, can  
 CC be used for the in vivo imaging of organs, receptors and receptor  
 CC containing tissues and/or atherosclerotic plaques, e.g. for the  
 CC diagnosis of breast or prostatic carcinoma. They can also be used  
 CC for cytostatic agent delivery.  
 CC  
 SO Sequence 21 AA;

Query Match 1.3%; Score 7; DB 18; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 1 CSCSSWL 7

RESULT 37  
 AAW14963 standard; peptide; 21 AA.

AAW14963;  
 XX  
 AC AAW14963;  
 XX  
 DT 28-NOV-1997 (first entry)  
 XX  
 DE Conjugating peptide for radio-therapeutic/diagnostic agent.  
 XX  
 KW Conjugate; diseased tissue; endothelin; angiotensin; chemotactic;  
 KW radiodiagnostic; radiotherapeutic; diagnosis; tumour; ischaemia;  
 KW atherosclerosis; vascular disorder; ischemia.  
 XX  
 OS Synthetic.  
 XX

Key Location/Qualifiers  
 FH Disulfide-bond 1..15  
 FT Disulfide-bond 3..11

XX W09710852-A2.  
 XX  
 PD 27-MAR-1997.  
 XX

PE 19-SEP-1996; 96WO-DE01821.  
 XX  
 PR 21-SEP-1995; 95DE-4036781.  
 XX

PA (UYBE-) UNIV BERLIN INST DIAGNOSTIKFORSCHUNG.  
 XX

PI Dinkelborg L, Erber S, Hilger CS, Kramp W, Platzeck J;  
 PI Raduechel B;  
 XX

DR WPI; 1997-202627/18.  
 XX

XX New sulphide-sulphonamide bi:functional chelating agents forming  
 PT new, stable technetium or rhenium complexes and conjugates useful as  
 PT radiodiagnostic and radio-therapeutic agents  
 XX

PS Claim 13; Page 43; 50pp; German.  
 XX

XX XSNS-type bifunctional sulphide containing sulphonamide chelating  
 CC agents and their technetium or rhenium complexes, can form  
 CC covalently bonded conjugates with substances selectively enriched  
 CC in diseased tissue, e.g. the present peptide. The substance is an  
 CC endothelin or angiotensin peptide, or a partial sequence, analogue,  
 CC derivative or antagonist, or a chemotactic peptide.  
 CC The conjugates are radiodiagnostic or radiotherapeutic agents,  
 CC useful in the diagnosis of tumours, ischaemia or atherosclerotic  
 CC vascular disorders.  
 CC  
 XX

SQ Sequence 21 AA;  
 XX

Query Match 1.3%; Score 7; DB 18; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 1 CSCSSWL 7

RESULT 38

AAW65855  
 ID AAW65855 standard; peptide; 21 AA.  
 XX  
 AC AAW65855;  
 XX

DT 19-OCT-1998 (first entry)  
 XX

DE Endothelin receptor binding peptide part of conjugate.  
 XX

KW endothelin receptor; antagonist; atherosclerosis; radionuclide;  
 KW tyrosine kinase blocker; chemotherapy; antibody; conjugate.  
 XX

OS Synthetic.  
 XX

PN DE19652374-A1.  
 XX

PD 10-JUN-1998.  
 XX

PF 04-DEC-1996; 96DE-1052374.  
 XX

PR 04-DEC-1996; 96DE-1052374.  
 XX

PA (SCHD) SCHERING AG.  
 XX

PI Blume F, Dinkelborg L, Hilger C, Speck U;  
 PI WPI; 1998-363559/32.  
 XX

DR WPI; 1998-363559/32.  
 XX

PT Therapeutic use, e.g. in treatment of atherosclerosis, of endothelin  
 PT conjugates - which comprise residue which can bind endothelin  
 PT receptor, conjugated to groups such as radionuclides or protein  
 PT tyrosine kinase blockers  
 XX

PS Claim 3; Page 15; 21pp; German.  
 XX

XX The invention relates to (A) the use of conjugates of the following  
 CC formula as therapeutic agents: E-Wn; where E = a residue which can bind  
 CC an endothelin receptor, and is derived from an endothelin, an endothelin  
 CC analogue, an endothelin derivative, an endothelin partial sequence or an  
 CC endothelin antagonist; W = an active group which: (i) is a radionuclide  
 CC or (ii) is derived from a chemotherapeutic agent, a complex with a  
 CC radioactive metal isotope, an antibody, an antibody fragment, a peptide,  
 CC a carbohydrate, an oligonucleotide, a protein tyrosine kinase blocker, an  
 CC anti-thrombotic agent, a coagulation cascade inhibitor, a hormone, growth  
 CC factor inhibitor, a medicament, a thrombocyte aggregation inhibitor, an  
 CC anti-inflammatory, a calcitonin antagonist, a lipid lowering agent or an  
 CC anti-proliferative agent; n = 1-100, especially 1-10, and (B) conjugates  
 CC of the above formula in which W is an active group which: (i) is a  
 CC radionuclide of the elements Al, Ba, Br, C, F, N, O or P or (ii) is  
 CC derived from a chemotherapeutic agent, an antibody, an antibody fragment,  
 CC a peptide, a carbohydrate, an oligonucleotide, a protein tyrosine kinase  
 CC blocker, an anti-thrombotic agent, a growth factor inhibitor, a  
 CC medicament, a hormone, a thrombocyte aggregation inhibitor, an anti-  
 CC inflammatory, a calcitonin antagonist, a lipid lowering agent or an anti-  
 CC proliferative agent. The conjugates may be used as therapeutic agents,  
 CC especially for treatment of cardiovascular disorders such as  
 CC atherosclerosis. They may be used in treatment of asthma, cerebral  
 CC infarction, subarachnoid haemorrhage, pre-eclampsia, renal disorders, etc.  
 CC The conjugates become enriched in cells in which endothelin receptors  
 CC are expressed. Even at low doses a therapeutically effective enrichment  
 CC of the active agent at desired sites can be achieved. Unbound conjugate  
 CC is rapidly eliminated from the body, reducing side effects. The present  
 CC sequence represents a specifically claimed endothelin receptor binding  
 CC peptide which forms part of the conjugate.  
 CC  
 XX

SQ Sequence 21 AA;  
 XX

Query Match 1.3%; Score 7; DB 19; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 |||||  
 1 CSCSSWL 7

Db 1 CSCSSWL 7

RESULT 39  
AAW65848  
ID AAW65848 standard; peptide; 21 AA.  
XX  
AC AAW65848;  
XX  
DT 19-OCT-1998 (first entry)  
XX  
DE Endothelin receptor binding peptide part of conjugate.  
XX  
KM endothelin receptor; antagonist; atherosclerosis; radionuclide;  
KW tyrosine kinase blocker; chemotherapy; antibody; conjugate; disulphide.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 1..15  
FT Disulfide-bond /label= disulphide\_bond  
FT Disulfide-bond 3..11  
FT Disulfide-bond /label= disulphide\_bond  
XX  
PN DEL9652374-A1.  
XX  
PD 10-JUN-1998.  
XX  
PE 04-DEC-1996; 96DE-1052374.  
XX  
PR 04-DEC-1996; 96DE-1052374.  
XX  
PA (SCHD ) SCHERING AG.  
XX  
PI Blume F, Dinkelborg L, Hilger C, Speck U;  
XX  
DR WPI; 1998-363559/32.  
XX  
PT Therapeutic use, e.g. in treatment of atherosclerosis, of endothelin  
PT conjugates - which comprise residue which can bind endothelin.  
PT receptor, conjugated to groups such as radionuclides or protein  
PT tyrosine kinase blockers  
XX  
PS Claim 3; Page 15; 21pp; German.

The invention relates to (A) the use of conjugates of the following formula as therapeutic agents: E-Wn; where E = a residue which can bind an endothelin receptor, and is derived from an endothelin, an endothelin analogue, an endothelin derivative, an endothelin partial sequence or an endothelin antagonist; W = an active group which: (i) is a radionuclide or (ii) is derived from a chemotherapeutic agent, a complex with a radioactive metal isotope, an antibody, an antibody fragment, a peptide, a carboxylate, an oligonucleotide, a protein tyrosine kinase blocker, an anti-thrombotic agent, a coagulation cascade inhibitor, a hormone, growth factor inhibitor, a medicament, a thrombocyte aggregation inhibitor, an anti-inflammatory, a calcium antagonist, a lipid lowering agent or an anti-proliferative agent; n = 1-100, especially 1-10, and (B) conjugates of the above formula in which W is an active group which: (i) is a radionuclide of the elements At, Ba, Br, C, F, N, O or P or (ii) is derived from a chemotherapeutic agent, an antibody, an antibody fragment, a peptide, a carboxylate, an oligonucleotide, a protein tyrosine kinase blocker, an anti-thrombotic agent, a growth factor inhibitor, a medicament, a hormone, a thrombocyte aggregation inhibitor, an anti-inflammatory, a calcium antagonist, a lipid lowering agent or an anti-proliferative agent. The conjugates may be used as therapeutic agents, especially for treatment of cardiovascular disorders such as atherosclerosis. They may be used in treatment of asthma, cerebral infarction, subarachnoid haemorrhage, preclampsia, renal disorders, etc. The conjugates become enriched in cells in which endothelin receptors are expressed. Even at low doses a therapeutically effective enrichment of the active agent at desired sites can be achieved. Unbound conjugate is rapidly eliminated from the body, reducing side effects. The present sequence represents a specifically claimed endothelin receptor binding

CC peptide which forms part of the conjugate.  
XX  
SQ Sequence 21 AA;  
XX  
Query Match 1.3%; Score 7; DB 19; Length 21;  
Best Local Similarity 100.0%; Pred. No. 30;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 32 CSCSSWL 38  
IIIIIIII  
Db 1 CSCSSWL 7

RESULT 40  
AAW65846  
ID AAW65846 standard; peptide; 21 AA.  
XX  
AC AAW65846;  
XX  
DT 19-OCT-1998 (first entry)  
XX  
DE Endothelin receptor binding peptide part of conjugate.  
XX  
KM endothelin receptor; antagonist; atherosclerosis; radionuclide;  
KW tyrosine kinase blocker; chemotherapy; antibody; conjugate; disulphide.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Disulfide-bond 1..15  
FT Disulfide-bond /label= disulphide\_bond  
FT Disulfide-bond 3..11  
FT Disulfide-bond /label= disulphide\_bond  
XX  
PN DEL9652374-A1.  
XX  
PD 10-JUN-1998.  
XX  
PE 04-DEC-1996; 96DE-1052374.  
XX  
PR 04-DEC-1996; 96DE-1052374.  
XX  
PA (SCHD ) SCHERING AG.  
XX  
PI Blume F, Dinkelborg L, Hilger C, Speck U;  
XX  
DR WPI; 1998-363559/32.  
XX  
PT Therapeutic use, e.g. in treatment of atherosclerosis, of endothelin  
PT conjugates - which comprise residue which can bind endothelin.  
PT receptor, conjugated to groups such as radionuclides or protein  
PT tyrosine kinase blockers  
XX  
PS Claim 3; Page 14; 21pp; German.

The invention relates to (A) the use of conjugates of the following formula as therapeutic agents: E-Wn; where E = a residue which can bind an endothelin receptor, and is derived from an endothelin, an endothelin analogue, an endothelin derivative, an endothelin partial sequence or an endothelin antagonist; W = an active group which: (i) is a radionuclide or (ii) is derived from a chemotherapeutic agent, a complex with a radioactive metal isotope, an antibody, an antibody fragment, a peptide, a carboxylate, an oligonucleotide, a protein tyrosine kinase blocker, an anti-thrombotic agent, a coagulation cascade inhibitor, a hormone, growth factor inhibitor, a medicament, a thrombocyte aggregation inhibitor, an anti-inflammatory, a calcium antagonist, a lipid lowering agent or an anti-proliferative agent; n = 1-100, especially 1-10, and (B) conjugates of the above formula in which W is an active group which: (i) is a radionuclide of the elements At, Ba, Br, C, F, N, O or P or (ii) is derived from a chemotherapeutic agent, an antibody, an antibody fragment, a peptide, a carboxylate, an oligonucleotide, a protein tyrosine kinase blocker, an anti-thrombotic agent, a growth factor inhibitor, a medicament, a hormone, a thrombocyte aggregation inhibitor, an anti-



CC inflammatory, a calcium antagonist, a lipid lowering agent or an anti-  
CC proliferative agent. The conjugates may be used as therapeutic agents,  
CC especially for treatment of cardiovascular disorders such as  
CC atherosclerosis. They may be used in treatment of asthma, cerebral  
CC infarction, subarachnoid hemorrhage, preclampsia, renal disorders, etc.  
CC The conjugates become enriched in cells in which endothelin receptors  
CC are expressed. Even at low doses a therapeutically effective enrichment  
CC of the active agent at desired sites can be achieved. Unbound conjugate  
CC is rapidly eliminated from the body, reducing side effects. The present  
CC sequence represents a specifically claimed endothelin receptor binding  
CC peptide which forms part of the conjugate.

XX  
SQ Sequence 21 AA;

Query March 1.3%; Score 7; DB 19; Length 21;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38

Db 1 CSCSSWL 7

Search completed: May 12, 2003, 10:07:34  
Job time : 82 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:49:16 ; Search time 27 Seconds  
(without alignments)  
806.484 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 2713

Sequence: 1 MRANPKTOAMPSETISLMKT.....KIMKRIAKILSLPIESL 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1115.5	41.1	493	YMDC_ECOLI	P75919 escherichia
2	632	23.3	502	Y190_HELPY	Q92mp2 helicobacte
3	628	23.1	502	Y190_HELPY	P56117 helicobacte
4	325	12.0	482	CUS2_BACSU	P71040 bacillus su
5	281	10.4	476	CUS_CLOPE	O9znc6 clostridium
6	276.5	10.2	503	CUS_BACFI	O66043 bacillus fi
7	274	10.1	503	CUS_BACHD	Q9x8z4 bacillus ha
8	270.5	10.0	500	CUSL_BACSU	P43660 bacillus su
9	259.5	9.6	481	CUS_PSEPU	P31048 pseudomonas
10	231	8.5	398	YMDJ_BACSU	P45865 bacillus su
11	222	8.2	413	YBHO_ECOLI	P75771 escherichia
12	212	7.8	486	CUS_ECOLI	P31071 escherichia
13	208	7.7	486	CUS_BUCAL	P53761 buchnera ap
14	169.5	6.2	455	PSS_HAEIN	P44704 haemophilus
15	151	5.6	373	YMAP_STRMU	P34001 streptococc
16	139.5	5.1	933	PID2_RAT	P70498 rattus norv
17	139	5.1	762	PDEL_ARATH	O9c888 arabidopsis
18	137.5	5.1	933	PID2_HUMAN	O14939 homo sapien
19	130	4.8	451	PSS_ECOLI	P23830 escherichia
20	128.5	4.7	1380	SP14_YEAST	P36126 saccharomyc
21	127	4.7	933	PID2_MOUSE	P79813 mus musculu
22	122	4.5	1036	PID1_CRIGR	O08684 cricetus
23	112	4.1	1203	ALAB_ARATH	O08685 arabidopsis
24	111.5	4.1	556	TYCC_SPRAT	O53728 streptomyce
25	111.5	4.1	6486	TYCC_BACBR	O30409 b tyrocidin
26	108.5	4.0	372	VENV_VACCC	P20638 vaccinia vi
27	108.5	4.0	372	VENV_VACCI	P26653 vaccinia vi
28	108.5	4.0	372	VENV_VACCP	P04021 vaccinia vi
29	106.5	3.9	372	VENV_VACCP	P29885 vaccinia vi
30	106.5	3.9	372	VENV_VACCP	P33815 variola vir
31	104.5	3.9	726	NUB4_YEAST	P52891 saccharomyc
32	104.5	3.9	1189	ALAB_ARATH	O91k90 arabidopsis
33	104	3.8	1096	PID1_ARATH	O91rt25 arabidopsis

34	103.5	3.8	1256	ATL_STNAU	P52081 staphylococ
35	103	3.8	1074	PLD1_MOUSE	O92280 mus musculu
36	102	3.8	820	PID1_ARATH	P58766 arabidopsis
37	101.5	3.7	673	FLUD_TREMA	O9kxw7 treponema m
38	101	3.7	812	PDAL_ORYSA	O43007 oryza sativ
39	100.5	3.7	438	YQPR_BACSU	P54475 bacillus su
40	100.5	3.7	485	C601_DROME	O9v979 drosophila
41	100.5	3.7	593	FRZE_MYXXA	P31759 myxococcus
42	100.5	3.7	618	GIDA_CAUCR	O9x8f8 caulobacter
43	100	3.7	832	DPRO1_TREAO	P19821 thermus aqu
44	99	3.6	1039	PID2_ARATH	O9m9w8 arabidopsis
45	98.5	3.6	747	YME1_YEAST	P32795 saccharomyc

## ALIGNMENTS

RESULT 1  
ID YMDC\_ECOLI STANDARD; PRT; 493 AA.  
AC P75919:  
DT 01-NOV-1997 (rel. 35, Created)  
DT 01-NOV-1997 (rel. 35, Last sequence update)  
DT 15-JUN-2002 (rel. 41, Last annotation update)  
DE Hypothetical protein ymdc.  
GN YMDC OR B1046.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97061202; PubMed=8905232;  
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,  
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
RA Yano M., Horiiuchi T.;  
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map.";  
RL DNA Res. 3:137-155(1996).  
CC -I- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
SYNTHASE SUBFAMILY. STRONG, TO H. PYLORI HP0190.  
CC -I- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: AE00206; AAC74130.1;  
DR EMBL: D90741; BAA35836.1;  
DR EMBL: D90742; BAA35844.1;  
DR ECGene: EGI3875; ymdc.  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC; 2.  
DR SMART: SM00155; PLDC; 2.  
DR PROSITE: PS50035; PLD; 2.

FW	Hypothetical protein: Transferase; Repeat; Complete proteome.	PLD PHOSPHODIESTERASE 1.
FT	DOMAIN 145 172	PLD PHOSPHODIESTERASE 2.
FT	DOMAIN 384 411	PLD PHOSPHODIESTERASE 2.
SO	SEQUENCE 493 AA; 55926 MW; 495604ADBD8492 CRC64;	
	Query Match	41.1%; Score 1115.5; DB 1; Length 493;
	Best Local Similarity	48.3%; Pred. No. 1.7e-72;
	Matches 224; Conservative 81; Mismatches 142; Indels 17; Gaps 5;	
QY	72 GLSDIYILNDPHEAARAAALLESHEHSDLOYIMRNDISGRLLFNLYVLAEGVRR 131	
Db	37 GQCGLEPLEKSLDAARARLEAMEHETLDVYIYIWMDSGRLLFSALLAAKGVRR 96	
QY	132 LLDDNNTRGDLDDLALDASHNPIEVRLEFPFLKKRWALGYLDFPRLNRMHKSFYA 191	
Db	97 LLLDNNTPRGDLIDLRLDLSHRIEVRLEFPFLRPGLRYGTDFSRNRMHKSFYV 156	
QY	192 DNRAIILIGRNGIDYEFYKGEDEVYADLDILATGSVGEVSHDEPDYMAASHANATRI 251	
Db	157 DGVVTLVGGRNIGDYFEGAGEPFLSDLDVMAIGVEVADAFARWYCKSVSPLOQVL 216	
QY	252 RSGNIGKG-----LQALGYNDEFSSRHALLRETFEVSPLYOKIOTGRIDMSQVTRL 304	
Db	217 ---DVEEGEMADRIELPASMWNDAHT---RYLKMESSPFINLVLVGTLEPLIAKIRL 269	
QY	305 ISDDPAKGLDRDRKPPPLAGLQDALKQPEKSVLYSPFYVTAKSGTDALAKLVODGIDV 364	
Db	270 LSDDPKAEGRKKRSLLPQRLFDIMGSPSEKIDIIISFYFTRAGVQDLRMVRRKGYKI 329	
QY	365 TVLTNSLOATDVAAVHSGVYKRYKRLPLAIGIKLYELOPHAVPAT-KKQGLRGSSVTSIH 423	
Db	330 AILTNLSLANDVAAVYAHAGYARAKRKLLKRGVLYELKPTREOSSTLHDGRTIGNSGASIH 389	
QY	424 AKTFIVDKRRLFIFGSNLDPRSARLNTENGVIYSEPKIAEQMERTLADTSPEYAVRYTL 483	
Db	390 AKTFESIDKRTVEFISFNDFPRSTLLNTENGVEFIESETLAQLDKRFIQSQYDAWQLRID 449	
QY	484 RHNRLQWHDPAATR-K--TYPNEPEAKLWRIAKIITSLPIESLL 525	
Db	450 RMGRINMWDRAAKKEITILKKPEATSGFMRKVRWVYLSILPEVWLL 493	
RESULT 2		
Y190_HELPJ	Y190_HELPJ	
ID	Y190_HELPJ	STANDARD: PRT: 502 AA.
AC	Q9ZMP2	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Hypothetical protein JHP0176.	
GN	JHP0176.	
OS	Helicobacter pylori J99 (Campylobacter pylori J99).	
OC	Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.	
CC	NCBI_TaxID=85963;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=99120557; PubMed=9923682;	
RA	Ahn R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G., Tummino P.J., Caruso A., Uria-Nikelsen M., Mills S.D., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F., Trust T.J.;	
RA	"Genomic sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori.";	
RT	Nature 397:176-180(1999).	
RL	-1- SUBCELLULAR LOCATION: Membrane-bound (potential).	
CC	-1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN SYNTHASE SUBFAMILY. STRONG, TO E-COLI YMDC.	
CC	-1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation -	

[illegible]

RP SEQUENCE FROM N.A.  
 RC STRAIN-26695 / ATCC 700392;  
 RX MEDLINE-97394467; PubMed-9252185;  
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,  
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
 RA Cotton M.D., Melman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,  
 RA Hayes W.S., Borodovsky M., Karpi P.D., Smith H.O., Fraser C.M.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the gastric pathogen *Helicobacter*  
 RT *pylori*.";  
 RL Nature 388:539-547(1997).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
 CC SYNTASE SUBFAMILY. STRONG, TO E.COLI YMDC.  
 CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE000539; AAD07257.1;  
 DR TIGR: HP0190;  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDc; 2.  
 DR SMART: SM00155; PLDc; 2.  
 DR PROSITE: PS50035; PLD; 2.  
 KW Hypothetical protein; Transferase; Transmembrane; Repeat;  
 KW Complete proteome.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT DOMAIN 162 189 PLD PHOSPHODIESTERASE 1.  
 FT DOMAIN 396 423 PLD PHOSPHODIESTERASE 2.  
 FT SEQUENCE 502 AA; 58287 MW; C80911F2FB59C714 CRC64;  
 SQ  
 Query Match 23.1%; Score 628; DB 1; Length 502;  
 Best Local Similarity 31.1%; Pred. No. 1.5e-37;  
 Matches 163; Conservative 93; Mismatches 170; Indels 98; Gaps 19;  
 QY 52 TSKPRLNDLIQIRHPHNGLSIDY-----LNDPHEAARAALIESAE 97  
 DB 26 SSSPLSYD-----PYTPPGSLYAEKLEKPNHSAAILLEDGDLHLHYGLIRMSQ 77  
 QY 98 HSLDQYIWRNDISGRLEFNLYLAERGVRLDLDNNTRGD---DELLALDSHP 153  
 DB 78 KSIDQTYIKNDSSQVIAKELNANANGVAVRILLDND---GLDSDSDIML-LNPFK 133  
 QY 154 NIEVLENFVLRKMPALCY---LNDPPLNRMNKSTFADNRATILCGRINIGDEYFV 210  
 DB 134 NIEVLENFVLRKMPALCY---LNDPPLNRMNKSTFADNRATILCGRINIGDEYFV 210  
 QY 211 GDDVFADDDIATGTSVGEVSHDPFRYVASHANATRLIRSGNIGKGLQALGYDETS 270  
 DB 193 DLDITNFDLDALEFGGVAKAKESFEFRYRFRHSIPVS-LLRT-----ISDP 309  
 QY 271 RHALLRYRETVQSPLYQRIQGRIDMOSVQRL-----ISDP 309  
 DB 235 -HRLRK-NNAKELAKLHEKIPISAEKDKNOFEKKVNDIFDFOKYQYPIYGNALFLADSP 292  
 QY 310 AGCLDRDRKPRPIAGRLQALAKQPEKSVYLVSPYPTSGSDALAKLVQDGDIVTLN 369  
 DB 293 KK-IDPIYVS-PIKIAFEKALNNAKDSVFIASSYFPGKMKKIFKNQJSGELMILTN 350  
 QY 370 SLQATDVAAVHSGYVYRKPLKAGIKYELQPNHNAVPTKDKGLGSSVTSIAHTFTY 429  
 DB 351 SLSSDIALVYVGAMERYKRLQVLRMGANYEIRNDFNRQIKGR---FSIKHSLHGKTYF 407

QY 430 DCKRIFGSEFNLDPRSARLNTGEMGVIESPKIAEQMERTLADTSPRYAVRVLDRNRRLQ 489  
 DB 408 DDLUTLLGSEFNIDPRASVINTESAVLFDNPSPFAKRRLSLAKDHA-QQSHLVYRRH-RVI 465  
 QY 490 W-----HDPATKRYTPNEPEAKLMKRIAKILSLPIESL 524  
 DB 466 WEAVEGILIH-----KT---SPDTSFRLIKEMSKVPEREL 502  
 RESULT 4  
 CLS2\_BACSU STANDARD; PRT; 482 AA.  
 ID P71040;  
 AC 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable cardiolipin synthetase 2 (EC 2.7.8.-) (Cardiolipin synthase  
 DE 2) (Cl synthase 2).  
 GN YWNE.  
 OS *Bacillus subtilis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE-98015417; PubMed-9353933;  
 RA Prescian E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,  
 RA Hullo M.-F., Lelong C., Schleich S., Sekowska A., Song B.H.,  
 RA Villani G., Kunst F., Danchin A., Glaser P.;  
 RT "The *Bacillus subtilis* genome from geroBC (311 degrees) to l1cR (334  
 RT degrees)."  
 RL Microbiology 143:3313-3328(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-168;  
 RX MEDLINE-98044033; PubMed-9384377;  
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Borries R., Boursier L., Brans A., Bifum M., Brignell S.C., Bron S.,  
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Deville K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
 RA Ertlan K.D., Erlington J., Fabret C., Ferrari E., Foulger D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Ghim S.Y., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,  
 RA Guisepi G., Guy B.J., Haga K., Haelech J., Harwood C.R., Hentut A.,  
 RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,  
 RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Melland R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portellel D., Porwollik S., Prescott A.M.,  
 RA Priesgen E., Puje C., Purnelle B., Rapoport G., Ray M., Reynolds S.,  
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serrier P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosoato V., Uchiyama S., Vandenbol M., Vanlier F., Vassarotti A.,  
 RA Viart A., Wampuit R., Wedler E., Wedler H., Welzenegger T.,  
 RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256(1997).  
 CC -1- FUNCTION: Catalyzes the reversible phosphatidyl group transfer  
 CC from one phosphatidylglycerol molecule to another to form  
 CC cardiolipin (CL) (diphosphatidylglycerol) and glycerol (by  
 CC similarity)  
 CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =  
 CC diphosphatidylglycerol + glycerol.

```

CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
CC SYNTHASE SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y08559; CAB69864.1; -
DR EMBL: Z99122; CAB15676.1; -
DR Subtilisin; B012483; YME.
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 2.
DR PROSITE: PS50035; PLD; 2.
KW Transferase; Phospholipid biosynthesis; Transmembrane;
KW Repeat; Complete proteome.
FT TRANSMEM 3 23 . POTENTIAL.
FT TRANSMEM 34 54 . POTENTIAL.
FT DOMAIN 217 244 . PLD PHOSPHODIESTERASE 1.
FT DOMAIN 395 422 . PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 222 222 . POTENTIAL.
FT ACT_SITE 400 400 . POTENTIAL.
SQ SEQUENCE 482 AA; 55865 MW; 8231FC35DB3CAC63 CRC64;

Query Match 12.08; Score 325; DB 1; Length 482;
Best Local Similarity 24.88; Pred. No. 6.9e-16;
Matches 124; Conservative 67; Mismatches 186; Indels 124; Gaps 15;

QY 43 ERTESRHFTSKPVLDN-----LQIRHTPTNGLSIDYILNDPHEAFARALIESA 96
DB 83 EDLEKQKQFNNRATFDDKDLIYLMINNHAFVTEFED-NSVDVITGGRKQFQLSDISKA 141
QY 97 EHSLDLYIYMRNDISGRLENNYLAERGVRRVLLDDNTRQLDILLALDSHPNIE 156
DB 142 KDHIHLQYIYKGDGLGRKLRLQKAKGQVQVRLVDELGSRLRKFFKELREAGGH 201
QY 157 VRLNPFVLRKMRALGYLIDPRLNRBMHNSFTADNRATILGSGNIGDEYF---KYGE 212
DB 202 VEVEFPKSLRPNTL-----RLNRYNRNRKLVIIIDGMTGVGGFVNGDEYLGLPKFG- 252
QY 213 DTFVADLDILATGVGEVSHDFRYMASSHANATRIIRS-----GNIGKGLQALG 264
DB 253 --YMRDTHIRLOGTAVNAIQTRFILDWQASHHHTLTYPNHPDYGPKGVNQMGLVTSG 310
QY 265 YNDESRHALLRYETVQSPLYOKIQTRIDWQSVQTRLSDDPAKGLDRDRKRPITAG 324
DB 311 PDSE-----WEQIKNGYI-----KMISN----- 328
QY 325 RLQDALKQPEKSVYLVSPYFPTKSGTDLAKLYODGIDVYLTNSLQATDVAAVHSGVY 384
DB 329 -----AKSILIQTYRFPIDASLDLALRLACISGIDVIMITPN--KPDHAFYVWATL 378
QY 385 KYRRRLKAGIRLYELQDNHNAVPATKDKGLTSSVTSLHAKTFIVDGKRIFIGSFNDLP 444
DB 379 SYIGGLLAKAGATVY-----IYDNGF-----IHAKTIYVDELASVGTANIDVR 421
QY 445 SARLNTENGAVYIESPKIAEQMERTIAD---TSPEYAVYVTLDRNRLQWHDPAATKTYPN 501
DB 422 SFRLEFVNAIYDITIKKLVSPFKEDLVSRKFTYEYIQR----- 464
QY 502 EPEAKLMKRIAKITLSL-PI 521
DB 465 -----PLWIRIKESVSRLLSPI 481

```

```

ID CUS_CLOPE STANDARD; PRT; 476 AA.
AC Q9ZNC6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cardiolipin synthetase (EC 2.7.8.-) (Cardiolipin synthase) (CL
DE synthase).
GN CUS OR CLSD OR CPEI430.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_Taxid=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCYC 8237;
RX MEDLINE=20049836; PubMed=10585141;
RA Koyama M., Katayama S., Kaji M., Taniguchi Y., Matsushita O.,
RA Minami J., Morita S., Okabe A.,
RT "A Clostridium perfringens hem gene cluster contains a cyst(B)
RT homologue that is involved in cobalamin biosynthesis.";
RL Microbiol. Immunol. 43:947-957(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater ";
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC -1- FUNCTION: Catalyzes the reversible phosphatidyl group transfer
CC from one phosphatidylglycerol molecule to another to form
CC cardiolipin (CL) (diphosphatidylglycerol) and glycerol (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol +
CC diphosphatidylglycerol + glycerol.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
CC SYNTHASE SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB017186; BAA74786.1; -
DR EMBL: AF003190; BAB81136.1; -
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 2.
DR PROSITE: PS50035; PLD; 2.
KW Transferase; Phospholipid biosynthesis; Transmembrane;
KW Repeat; Complete proteome.
FT TRANSMEM 2 22 . POTENTIAL.
FT TRANSMEM 31 51 . POTENTIAL.
FT DOMAIN 207 234 . PLD PHOSPHODIESTERASE 1.
FT DOMAIN 389 416 . PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 212 212 . POTENTIAL.
FT ACT_SITE 394 394 . POTENTIAL.
FT CONFLICT 469 469 T -> I (in Ref. 1).
SQ SEQUENCE 476 AA; 55040 MW; 8BC32FC113F0A9CB CRC64;

Query Match 10.48; Score 281; DB 1; Length 476;
Best Local Similarity 24.58; Pred. No. 9.6e-13;
Matches 109; Conservative 79; Mismatches 166; Indels 90; Gaps 20;

QY 41 LEERTSRHFTSKRP-VRLDNI-----LQIRHTPTNGL-----SDIYILNDPHEAFAR 89
DB 66 LDDTKKQKYLNSFKSHYKLDNISLKYKDLIMNPNRNDNSTYTORNDIDLTFDANSLEFEM 125

```

QY 90 AALLESASHDLQYVIRNRDINSGRLFNELVYLAERGVRLDLD--NNTPGLD-DLL 146  
126 IDEINKAEKFLHMEGYIFKSDIEGKIIQALTKRAKEVEYKLLVDISGNSIHKKDDIKL 185  
QY 147 LALDSHPNIEVRLFNPFVLRKWRALGYLTDPRLNRHMKNSFTADNRATILGGRNIDE 206  
Db 186 KAAGG----DEKIFEPGFC-----YTN--LRIRYRHRKILILDSKVAFGLGNIGDE 233  
QY 207 YEKYGEDVFPVDDLILAGSVYGEVSHDFDYKWSHSHNNATRIIRSGNT-GKGLALGY 265  
Db 234 Y--LGKDK-----NIGH-----W--RDYTRIKGLAINDLEGKFLDMSY 269  
QY 266 NDEFSRHA-LIRYRETVESQPLYOIKQIGRIDMOSVQRLISDDPAKGLDNRKRPYAG 324  
Db 270 ANESDLDLDLKKYFLINPSTDLPKI-----IGAQIVSSGP-----DHEEQILKN 314  
QY 325 RLQDALQPEKSVYLVSPFVPTKSGTDALAKLVODGIDVTYLTNSLATDVAAVHSGV 384  
Db 315 GYFKFINSKAKNLTQFPYFVPEDEPMLEALRLAALSGVDVKIMLPG----NPDHKFMGI 370  
QY 385 --KRRKLLKGLITLYELQPMHVAVPATDKGLTSGSVSLAKFTYDGKRIFGSFMD 442  
Db 371 ANSFESLLNKGAVIY-----LYEKG-----LHAKTIVADSSICSVGTANMD 413  
QY 443 PRSARLNTMGVIVTESPKIAEQE 466  
Db 414 IRFSPLNESNIFITYNEAISKSM 437  
RESULT 6  
CLS\_BACFI  
ID CLS\_BACFI STANDARD; PRT; 503 AA.  
AC 066043;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Cardiolipin synthetase (EC 2.7.8.-) (Cardiolipin synthase) (CL synthase).  
GN CLS.  
OS Bacillus firmus.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1399;  
RN [1]  
RX MEDLINE=98103838; PubMed=9443601;  
RA Guo D., Tropp B.E.;  
RT "Cloning of the Bacillus firmus OF4 cls gene and characterization of its gene product."  
RL Blochm. Biophys. Acta 1389:34-42(1998).  
CC -1- FUNCTION: Catalyzes the reversible phosphatidyl group transfer from one phosphatidylglycerol molecule to another to form cardiolipin (CL) (diphosphatidylglycerol) and glycerol.  
CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =  
CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN SYNTHASE SUBFAMILY  
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>).  
CC  
DR EMBL: U88888; AAC05444.1; -  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDc: 2.  
DR SMART: SM00155; PLDc: 2.  
DR

DR PROSITE: PS50035; PUF; 2.  
 KM Transferase; Phospholipid biosynthesis; Transmembrane; Repeat.  
 FT TRANSMEM 5 25 POTENTIAL.  
 FT TRANSMEM 30 30 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT DOMAIN 238 265 POTENTIAL.  
 FT DOMAIN 416 443 PUF PHOSPHODIESTERASE 1.  
 FT ACT\_SITE 243 243 POTENTIAL.  
 FT ACT\_SITE 421 421 POTENTIAL.  
 SQ SEQUENCE 503 AA; 57907 MW; E02447184E987AF3 CRC64;  
 Query Match 10.2%; Score 276.5; DB 1; Length 503;  
 Best local similarity 24.4%; Pred. 2.2e-12;  
 Matches 107; Conservative 61; Mismatches 170; Indels 101; Gaps 13.  
 QY 74 SDIYLLDPHEFAARAALIESAEHSLDQYTWBNDSIGRLLEMLVYLAERGVYRLL 133  
 Db 140 SETKVLDTGKEFYAHIIQLAKMAEHIIHELYIVRHDDLGNIKIDLSKAKEGVHVRLL 199  
 QY 134 LDDNNTKRLDLLLLALDSHPNIEVYLFNPFVLRKRALCYLTDPRLRNRINKSFADN 193  
 Db 200 YDVGSMWLSKSYVELRDAGVEMVSFSPV-----KLPLTH--TINRNRKTIYIDG 251  
 QY 194 RATILGKRNIGDEYKVEDFYFA---DIDILATGSVGEVSHDDPRW---ASHAHNA 247  
 Db 252 VVGFVGGIATIGDEY--LGKDAIFGWRDTHLYVREAVRTLLDILDMHYTGGETILNQ 309  
 QY 248 TRIIRSGNIGK---GLQALGYNDTSRHALLRYRETEQSPLYQKIQTGRIQMSQVTRL 304  
 Db 310 TYLSPSLSMFKGDDGVQMIASGPT-----RMEVKKLFFSMITSK----- 351  
 QY 305 ISDDPAKGLDRKRPRPIAGRLQDLAKQPEKSVLYSPFVYTKSGTDALAKLYDDGIDY 364  
 Db 352 -----KSIWIASPFPIRDDILSLKTAALSGIDY 381  
 QY 365 TYLTNSLGATDVAAYHSGVYKRPDLKAGIKLYELDPRHNAVPAUKDKGLTGSVTSLSA 424  
 Db 382 RILVFN--RPDKRIVFHASRSTFPELLEAGVAYEY-----NRGF-----MHS 422  
 QY 425 KTFIVDGKRIEFGSENLDPKRSARLNTENGVIYIESPKIAEQMERTLADTSPRAYRYTLDR 484  
 Db 423. KIIVDHEIASIGTSMQMRSPFLNMFVNA-----YLRTSSVFLVSDYDYDLEH 473  
 QY 485 HNRLOWHDPATRKTYPNED 503  
 Db 474 SNOINF-----SLEKKNRP 486  
 RESULT 7  
 ID CLS\_BACHD STANDARD: PRT: 503 AA.  
 AC O9K824;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Cardiolipin synthetase (EC 2.7.8.-) (Cardiolipin synthase) (CL synthase).  
 DE CLS OR BH2858.  
 GN Bacillus halodurans.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=86665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N., Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kunara S., Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -!- FUNCTION: Catalyzes the reversible phosphatidyl group transfer from one phosphatidylglycerol molecule to another to form

CC cardiolipin (CL) (diphosphatidylglycerol) and glycerol (By  
CC similarity).

CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =  
CC diphosphatidylglycerol + glycerol.

CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).

CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
CC SYNTHASE SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AP001516; BAB06577.1; -  
DR InterPro; IPR001736; PLD.  
DR Pfam; PF00614; PLDC; 2.  
DR SMART; SM00155; PLDC; 2.  
DR PROSITE; PS00035; PLD; 2.  
KW Transferase; Phospholipid biosynthesis; Transmembrane;  
KW Repeat; Complete proteome.  
FT TRANSMEM 7 25 POTENTIAL.  
FT TRANSMEM 29 51 POTENTIAL.  
FT TRANSMEM 58 80 POTENTIAL.  
FT DOMAIN 238 265 PLD PHOSPHODIESTERASE 1.  
FT DOMAIN 416 443 PLD PHOSPHODIESTERASE 2.  
FT ACT\_SITE 243 243 POTENTIAL.  
FT ACT\_SITE 421 421 POTENTIAL.  
SO SEQUENCE 503 AA; 58101 MW; 2505BE2B7A5821C8 CRC64;

Query Match 10.1%; Score 274; DB 1; Length 503;  
Best Local Similarity 24.3%; Pred. No. 33e-12;  
Matches 109; Conservative 58; Mismatches 188; Indels 94; Gaps 12;

OY 61 ILQIRHTPTNLS---DIYLLNDHEAFARAALIESHSLDIQYTWNDISGRLLF 117  
DB 124 LFLRLHRLANNPISSTNRYVLTDEKETFHAKQALRMATHTIHIEYIVRDEIGQETK 183  
OY 118 NLYVLAENGAVRLLLDNNRTGDLDDLALDSHPNIEVRLENFVLRKRWALGYLDFE 177  
DB 184 ELMQKAKEGHVRFLYDGVGSMKLSKSTIQDLKQAGVIVFAFYKL----- 231  
OY 178 P-----RLNRHMNKSFTADNRATILGKRNIGDYFKVGEIVFA---DDILATGSVGE 230  
DB 232 PFINHTIVRNHRKLIIVDGTGVFGGLNIGDEY--LCKDPYFGFWROTHLVYRGCAVRT 289  
OY 231 VSHDDRIVWASHSAHNATRIIRSGNIGKQLQALGYNDETSRHALLRYETVEDOSPLYOKI 290  
DB 290 LQILFLRWMAHETG-----ETIIKPSYLSPA 315  
OY 291 QGRIDMOSYQRLISDPAKGLDRRRKPIAGRLQDALKQPEKSVYLVSPYPTKSG 350  
DB 316 LTNMKDGGVQ--MTASGP-----DTRWEINKKLFESMTSAKSIWTSPIFIPDEI 367  
OY 351 TDALAKIVODGIDVTLVNTSLQATDVAAVHSGYVYRKPLKAGIKLYELQPNHVAIPATK 410  
DB 368 LSAKIALAGIDVRLIVN--RPRKRIYFHSRSYFPFLLEAGYKVIY----- 415  
OY 411 DKGLTGSSVTSIAAKTFIVDGKRIFGSFNDPRSARLTENGVAIESPKIAEQNERTILA 470  
DB 416 TRGF-----LHKKIIVDNEIASIGTSNMDRSPLNFVEVNAF-----LYRTKS 459  
OY 471 DTSPEYAVRVLDRNRRLQMHDPATRKTY 499  
DB 460 VTTLVSDPYVDLEHTNQIRFEQFRNRAWY 488

RESULT 8  
CLSI\_BACSU  
ID CLSI\_BACSU STANDARD; PRT; 500 AA.

AC P45860;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable cardiolipin synthetase 1 (EC 2.7.8.-) (Cardiolipin synthase  
DE 1) (CL synthase 1).  
GN YWIE.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98015417; PubMed=9353533;  
RA Preecean E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,  
RA Hullo M.-F., Lelong C., Schleich S., Sekowska A., Song B.H.,  
RA Villani G., Kunst F., Danchin A., Glaser P.;  
RT "The Bacillus subtilis genome from gerBC (311 degrees) to llerC (334  
RT degrees)".  
RL Microbiology 143:3313-3328(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertoletti M.G., Bessieres P., Bilotin A., Borchert S.,  
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Ertlan K.D., Errington J., Fabre C., Ferrari E., Foulger D.,  
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porrolik S., Prescott A.M.,  
RA Preecean E., Pujić P., Purrelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takenuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Wellzenegger T.,  
RA Winters P., Wipal A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis".  
RL Nature 390:249-256(1997).  
CC -1- FUNCTION: Catalyzes the reversible phosphatidyl group transfer  
CC from one phosphatidylglycerol molecule to another to form  
CC cardiolipin (CL) (diphosphatidylglycerol) and glycerol (By  
CC similarity).

CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =  
CC diphosphatidylglycerol + glycerol.

CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).

CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
CC SYNTHASE SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



```

CC -----
DR EMBL: 249884; CAA90049.1; -
DR EMBL: 249782; CAA89861.1; -
DR EMBL: 269123; CAB15752.1; -
DR Subtilisin; B011305; ywie.
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 2.
DR PROSITE: PS50035; PLD; 2.
DR Transferrase: Phospholipid biosynthesis; Transmembrane;
KW Repeat: Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT DOMAIN 237 264 PLD PHOSPHODIESTERASE 1.
FT DOMAIN 413 440 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 242 242 POTENTIAL.
FT ACT_SITE 418 418 POTENTIAL.
SQ SEQUENCE 500 AA; 58247 MW; 3BEF85FDAF0681B8 CRC64;

Query Match
Best Local Similarity 10.0%; Score 270.5; DB 1; Length 500;
Matches 100; Conservative 63; Mismatches 175; Indels 111; Gaps 13;

QY 48 RHENSKPYRLDNIQIRHTPTHTGLSD-----IYLLNDPHEA 85
D 94 RMYNREK---LRKLFKEETPEVTGLKQDERFFYSIRAHMNIKTSNIKVLKNGEET 150
QY 86 FARALALISAEHSDLOYIMRNDISGRLEFNLVLAERGVRVRLDDNNTRGDLDL 145
D 151 FPDIFKAMKRAESYIHIEYMFSDMLGQMDIMMEKAGVEVRFVLDAGASMKIARR 210
QY 146 LIALDSHPNIEVLFNPFLRKWRALGYLTDPRLRMRMKNKSFADNRATILGKRNIGD 205
D 211 DIMRMQAGVDIYFESPL-----KYGFEHQ--KLNFRNRKVIIDGKGFVGLNWK 252
QY 206 EYFKVEDTVFADLDILATGVSVEGSHD--FDRYWASHSAHNATRIIRSGNICKGLQALG 264
D 263 EY-----ISRDPIYIGFWRD-----THRLGEIYQIHLAIF 293
QY 265 YND--ETSRHALLRRE-----TVEOSPLYOKIOTGRIDMOSVOTRLISDDPAKGLDRDR 317
D 294 MLDMEVYSNEVLIDQEEYNTPVPEGGGIQIYATG-----PD 331
QY 318 RRPPIAGRDALQKPEKSVYLVSPEVPTKSGTDALAKLVODGIDVTLNLSQATDVA 377
D 332 MKESMDLYEMKISSAKQSIWITPTFVFNESIRTKAKAATGVEYRWVP--EKNDSE 389
QY 378 AVHSGYVYTRKPLKAGIKLYELQPNHNAVPAIKDKGLTGSSVTSLSHAKTFIVDGKRLFIG 437
D 390 LQYASRSYFPELLLEGIEVSYQ-----KGF-----MHQKWIIDGLASV 432
QY 438 SFNLDPRASRLNTEMGVIESPKIAEQME 466
D 433 TANMDRSQFLNEVNVFETDAEAIPTLE 461

RESULT 9
CLS_PSEPU STANDARD: PRT: 481 AA.
AC P31048;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable cardiolipin synthetase (EC 2.7.8.-) (Cardiolipin synthase)
DE (CL synthase).
GN CLS.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN (1)
RP SEQUENCE FROM N.A.

```

```

RC STRAIN-PPG2;
RA Lorenz D., Sokatch J.R.;
RL Submitted (Aug-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the reversible phosphatidyl group transfer
CC from one phosphatidylglycerol molecule to another to form
CC cardiolipin (Cl) (diphosphatidylglycerol) and glycerol (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =
CC diphosphatidylglycerol + glycerol.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
CC SYNTHASE SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: X55704; CAA39233.1; -
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 2.
DR PROSITE: PS50035; PLD; 2.
KW Transferrase: Phospholipid biosynthesis; Transmembrane; Repeat.
FT TRANSMEM 9 32 POTENTIAL.
FT TRANSMEM 39 62 POTENTIAL.
FT DOMAIN 220 247 PLD PHOSPHODIESTERASE 1.
FT DOMAIN 394 421 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 225 225 POTENTIAL.
FT ACT_SITE 399 399 POTENTIAL.
SQ SEQUENCE 481 AA; 54328 MW; B23A61677AE6B93F CRC64;

Query Match
Best Local Similarity 9.6%; Score 259.5; DB 1; Length 481;
Matches 116; Conservative 75; Mismatches 191; Indels 115; Gaps 16;

QY 36 SWLPLEERTESKHFNTSKPYRLDNIQIRHTPTHTGL-----SDIYLLNDPHEAFARA 90
D 87 NMRPWEAEALTAR-----ESESVALRLAMKRLGMPCLANNQVLLVNGKATPDAIF 138
QY 91 ALLESASHSDLOYIMRNDISGRLEFNLVLAERGVRVRLDDNNTRGDLDLALD 150
D 139 AALEKARDVLYVQFTIHDITGLKALDQLLKRAABGVQVFLYDRYGSALPASISQOL 198
QY 151 SHPNIEVRLNPEVLRKWRALGYLTDPRLRMRMKNKSFADNRATILGKRNIGDEYFKV 210
D 199 RNCGVQI--HAFATRR-----GWENNR--QVNFNRKIVVVDGLGFIGSHNVGDEV--L 248
QY 211 G--EDTVFADLDILATGVSVEGSHD--FDRYWASHSAHNATRIIRSGNICKGLQALG 265
D 249 GGHPRQSPWRDTHVOISGPLACIOESFAEDWYAT-----ROLPELIL 292
QY 266 NDETSRALLRRETVYQSPLYOKIOTGRIDMOSVOTRLISDDPAKGLDRKRPPIAGR 325
D 293 PDAYP-----DNGVLCQALASGPRADPQETCSLF----- 320
QY 326 LDALQKPEKSVYLVSPEVPTKSGTDALAKLVODGIDVTLNLSQATDVAHVHSGYVK 385
D 321 FLEAHISATRRWVITSPYFIPDEAVFALRLAVLRGDVAVLLPS--RPDRHLYVAASSL 378
QY 386 YRKPLKAGIKLYELQPNHNAVPAIKDKGLTGSSVTSLSHAKTFIVDGKRLFIG 445
D 379 FAEAVRAGVRMRYPQGF-----LHQKVVVLDVEVSAIGSANDNRS 421
QY 446 ARLNTEGVVIESPKIAEQMERTLA---DTSPEYAVVTLDRNRRLQMDHPRARKYTPNE 502
D 422 FLNFETITLTVDRNPAQOVHMLIKDFQAREITAEEDSRTH-RLO----- 467
QY 503 PEAKLWKRIAKTILSL 519

```

DB 468 ---OLGMRIARLSPITL 481

RESULT 10  
YWJE\_BACSU STANDARD: PRT: 398 AA.

AC P45865: 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ywje.  
GN YWJE.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
NX NCBI\_TaxID=1423;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98015417; PubMed=9353933;  
RA Prescan E., Moszer I., Boursier L., Cruz Ramos H.C., De La Fuente V.,  
RA Hullo M.-F., Lelong C., Schleich S., Sekowska A., Song B. H.,  
RA Villani G., Kunst F., Danchin A., Glaser P.;  
RT "The Bacillus subtilis genome from gerBC (311 degrees) to llyC (334  
RT degrees)".  
RL Microbiology 143:3313-3328(1997).  
RN 12  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azavedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Boursier L., Boursier L., Brans A., Braun M., Brigneau S.C., Bron S.,  
RA Brouillet S., Bursch J., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codan J.J., Conerton I.F., Cummings N.J., Daniel R.A.,  
RA Delizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Entian K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Gilm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,  
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Hentut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Katamata D., Kasahara Y., Kleatr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauei C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Patro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M.,  
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Setiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
RA Sotokun A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
RA Totsu V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the Gram-positive bacterium Bacillus  
RT subtilis".  
RL Nature 390:249-256(1997).  
RC -1 SUBCELLULAR LOCATION: Membrane-bound (potential).  
CC -1 SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
CC SYNTHASE SUBFAMILY.  
CC -1 SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: 249782; CAA89866.1;  
DR EMBL: 299123; CAB15747.1;  
DR Subtilist: BG11310; ywje.  
DR Interpro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC; 2.  
DR SMART: SM00155; PLDC; 2.  
DR PROSITE: PSS0035; PLD; 2.  
KW Hypothetical protein; Transferase; Transmembrane; Complete proteome;  
KW Repeat.  
FT TRANSMEM 3 23 POTENTIAL.  
FT DOMAIN 141 168 PLD PHOSPHODIESTERASE 1.  
FT DOMAIN 311 338 PLD PHOSPHODIESTERASE 2.  
SQ SEQUENCE 398 AA; 45842 MW; 4E233A4B9727756 CRC64;  
Query Match 8.5%; Score 231; DB 1; Length 398;  
Best Local Similarity 21.6%; Pred. No. 2.8e-09;  
Matches 94; Conservative 65; Mismatches 169; Indels 108; Gaps 13;  
OY 74 SDIYLLNDPHEAFARALIESAHSLSLQYIWRNDISGRLLFNLVLAERGVYRLL 133  
DB 43 SDIELHCGADLVERMMNDIROAASVHMFFLMKNDVSHMYTLTKKAQAGSYLL 102  
OY 134 LDDNNTRGDILLALDSPNIEVRLFN----PVLTKKRALGYLTPRLNRKHNKSF 189  
DB 103 LWMAGCRAIKTKALQTMKNAGVHVHMNRPREF-----FFHMOKRHNKRT 150  
OY 190 TADNRATILGHNIGDEYKVGEDYFADLDITATGSVYGEVSHDPFRWASHANMR 249  
DB 151 VLDKRGYIGGFNIAEY--LKKAKFGWEDYHL-RMIGEVHDLQTLFASDLKRN-- 205  
OY 250 IIRSGNIGKGLDALGYNDSTSHALLRRETVESQSPLYOKIOTGRI-----DMOSVQT 302  
DB 206 -----GIE-LG-----SDVWPKLOQGTISRKYATDGSLEN 236  
OY 303 RLISDPKAGLDRDRKRPPIAGRLQDALKQPKSVYIVSPYVPRKSGDALKVQGI 362  
DB 237 IYLAN-----IAQAKRLTVCPIYIPISKPLQDALINARKGV 274  
OY 363 DVTVLNLSIATDVAHVSGYKRPPLKAGIKLYELOPNHVPATDKGLTGSVTS 422  
DB 275 SVRIIVP--MKSDHPLVRAFAFYSELLDACCLIVRYQGF-----Y 315  
OY 423 HAKTIVQCKRFFTSFNIDPRASALNTEMGVVIS-----PKIAQMRITLADSP 474  
DB 316 HIKALILIDHISTIGANDKRSFLNEVNEIDAEFTSEVATIEDMKSELTWE 375  
OY 475 EYAVRVTLDRHNRLOM 490  
DB 376 DFKRTF--RQRPAEM 389  
RESULT 11  
YBHO\_ECOLI STANDARD: PRT: 413 AA.  
AC P75771:  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical protein ybho.  
GN YBHO OR B0789 OR Z1008 OR ECS0867.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562, 83334;  
RN 11  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,



	DR	EMBL:	U24198:	AAB60096.1:	-
	DR	EMBL:	U24199:	AAB60104.1:	-
	DR	EMBL:	U24200:	AAB60112.1:	-
	DR	EMBL:	U24201:	AAB60120.1:	-
	DR	EMBL:	U24202:	AAB60128.1:	-
	DR	EMBL:	U24203:	AAB60136.1:	-
	DR	EMBL:	U24204:	AAB60144.1:	-
	DR	EMBL:	U24205:	AAB60152.1:	-
	DR	EMBL:	AE000223:	AAC74331.1:	-
	DR	EMBL:	D90763:	BAA14781.1:	-
	DR	Ecogene:	Eg1608:	CIS.	
	DR	InterPro:	IPIR001736:	PLD.	
	DR	Pfam:	PF00614:	PLDC; 2.	
	DR	SMART:	SM00155:	PLDC; 2.	
	DR	PROSITE:	PS50035:	PLD; 2.	
	KM	Transferrase:	Phospholipid biosynthesis;	Transmembrane;	
	KW	Repeat:	Complete proteome.		
	FT	TRANSEM	7	POTENTIAL.	
	FT	TRANSEM	37	POTENTIAL.	
	FT	DOMAIN	219	PLD PHOSPHODIESTERASE 1.	
	FT	DOMAIN	399	PLD PHOSPHODIESTERASE 2.	
	FT	ACT SITE	224	POTENTIAL.	
	FT	ACT SITE	404	POTENTIAL.	
	FT	VARIANT	113	I -> S (IN STRAIN ECOR 60).	
	FT	VARIANT	283	E -> G (IN STRAIN ECOR 50).	
	FT	VARIANT	378	S -> S (IN STRAIN ECOR 16).	
	FT	VARIANT	426	S -> R (IN STRAIN ECOR 46).	
	SO	SEQUENCE	486 AA; 54822 MW; C8BDD69A5AA37DB CRC64;		
		Query Match	7.8%; Score 212; DB 1: Length 486;		
		Best Local Similarity	20.6%; Pred. No. 8.6e-08;		
		Matches 114; Conservative	76; Mismatches 207; Indels 156; Gaps 20.		
OY	17	LMKTRSLISLCLLC-----SCSSWLPLEE	43		
Db	28	LKKRAVPSPAMALLIYLPLVGIAYLVAGELHGLKRRARBARAMPSTAKTWLDLK-	86		
OY	44	FRESHF---TTSKPVRDLNLOIRHPHTNGL-----SDIYLNDPHEANAALIE	94		
Db	87	--ACHHFAEEENSSVAAPLFKLCEPR--OGIGVGKNQLOLTESDVMOALIRIDQ	139		
OY	95	SAESHLDLYTYWRNDISGRLLFNLYVLAERGVRRLDDNNRTGL----DDL--L	147		
Db	140	LARNHIEMEVFIWOGGMADQVAFSLMAARGIHCHRLMLDSAGSVAFRSPPELMRNA	199		
OY	148	ALDSHPNIEVFLFNPVLKRRALGYLDPPLNRMINKSTADRATILGGRINGD-E	206		
Db	200	GIEVEALKVNIMRFEL-----RRMDLRHKMTIMINDNYIAVTGSMMWVDP	246		
OY	207	YRK-----GGEIVFPDLDIATGSVVGESHDFRYTMASHAHNRRIIRSGNIKGLOA	262		
Db	247	YFKOAGVGO--WIDLAMAREGPATTAAGIIYSCDMEJETKRILLPREPDVINMFEDA	303		
OY	263	LGYNDETSHALLRRETEVOESPLOYKIQTGRIDMQSVOTRLISDDPAKGLDRDRRPPI	322		
Db	304	SGHTIHT-----IASGP-GPEDL----	322		
OY	323	AGRLODALKOPKSYLVLSPTVPPTKSGDIALAKLVODSIDVTYLTNSIQATDYAAVHS	382		
Db	323	HQALLTAAYSAREVLIIMTPPYFPPSDDELHAICTAORGVDSITLP--RKNDSSLVYWA	380		
OY	383	YVKRKPLLIKIKIYVELQPNHAVPATKDKGTGSSVSLSNAKTFFYDGKRFISFND	442		
Db	381	SNAFFTELLAGVKTIYO-----EGTL-----LTKSKVLVBGELSIVGTVND	423		
OY	443	PSASRLNTMGVAVIESPKIAEQMERTLADTSEPAYRVYTLDRHNRLQMHDPARTRTYPNE	502		
Db	424	MRSLLNLNEFITLAIDKGFAD-----LAAVDODYISRSL-----LDARLMLKR	468		
OY	503	PEAKIMKRIAIAKI 515			
Db	469	P---LMQVAERL 478			

```

RESULT 13
CLS_BUCAI
ID CLS_BUCAI STANDARD: PRT: 486 AA.
AC P57361:
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Cardiolipin synthetase (EC 2.7.8.-) (Cardiolipin synthase) (CL
  synthase).
GN CLS OR BU273.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
  symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_Taxid=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Tokyo 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
  Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -1- FUNCTION: Catalyzes the reversible phosphatidyl group transfer
  from one phosphatidylglycerol molecule to another to form
  cardiolipin (CL) (diphosphatidylglycerol) and glycerol (By
  similarity).
CC -1- CATALYTIC ACTIVITY: 2 Phosphatidylglycerol =
  diphosphatidylglycerol + glycerol.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
  SYNTHASE SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001118; BAB12983.1; -.
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 2.
DR PROSITE: PS50035; PLD; 2.
KW Transferrase; Phospholipid biosynthesis; Transmembrane;
  Repeat; Complete proteome.
FT TRANSMEM 3 23
FT TRANSMEM 3 23
FT DOMAIN 38 58 POTENTIAL.
FT DOMAIN 219 246 PLD PHOSPHODIESTERASE 1.
FT ACT_SITE 399 426 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 224 224 POTENTIAL.
FT ACT_SITE 404 404 POTENTIAL.
SQ SEQUENCE 486 AA: 56329 MW: 50A59BED7F6ABA56 CRC64;

Query Match 7.7%; Score 208; DB 1; Length 486;
Best Local Similarity 21.4%; Pred. No. 1.7e-07;
Matches 105; Conservative 81; Mismatches 163; Indels 142; Gaps 21;

OY 49 HNTSKPVRL-----DNIIQIRTPHTNGSLDIYLLNDPHEAFARAALIESAHSIDLOY 104
DB 114 HGIKSKIKILNTTKIMQIL-----IRDIYL-----ARKNIEMVF 149
OY 105 YLMR-----NDISGRLEFLVLAERGVRLLDNDNTRGIDLLALDSHPNIEVRL 159
DB 150 YLMKPGMADVADVAALID-----SARIGIHCRLMDSAGS-----IEFQSPWEIMR 197
OY 160 FNPFLKRRALGYLTFD--PRLNRRMHNSFTADNRATILGCGNIDGYF--KVGEDYVF 216
DB 198 KSGIQVVELAKYNLRFVRLRDYVRHKKILIDNTYIAISGSNLYDPYLFKSSSEIGOM 257

```

```

OY 217 ADDILATGSGVGEVSHDFEDRWASHAHNATRIIRSGNIGKGLQALGYNDETSRHALL 276
DB 258 IDLMTRESGPIATMTWIGIYSCDW-----ELEFGKIL----- 289
OY 277 YRETEVQSPLYQKIQTGRIDMOSVQTRLSDDPAKGLDRDRKRPPIAGRLQDAKQPEKS 336
DB 290 -----POLPKRKMILEMOSNKNASIQ--VIASGPG-----FLKNMIIQALTAIYSKRE 336
OY 337 VVLVSPFYVETKSGTGLATLAKLVODGIDYVTL-----TNSLOATDVAAVHSVYVYRKLK 392
DB 337 LIITTPYLPVSEDLLEICTAAGRGVEVSTITPLVNDSTILVKNASRY-----FFSELL 390
OY 393 AGIKLYELOPNHVPATKDKGLTGSVTSLSAAKTFYDGKRIFGSEFNLDPRSARLNTTEM 452
DB 434 TLVIDSDDFGRNLFQIONKYSISQO-----LIDKK--AW-----SMRAY 471
OY 453 GVIESPKIAEQM-----ERLATDSEYAVRVTLDRNRNLQMHDPATKTYPNPEPAKLW 508
DB 472 KRIAKTILSL 519
OY 509 KRIAKTILSL 519
DB 472 KRIAKTILSL 482

RESULT 14
PSS_HAEIN
ID PSS_HAEIN STANDARD: PRT: 455 AA.
AC P44704;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE CDP-diacylglycerol--serine O-phosphatidyltransferase (EC 2.7.8.8)
  (phosphatidylserine synthase).
GN PSSA OR H10425.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
  Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
  Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
  Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
  Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
  Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd."
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: CDP-diacylglycerol + L-serine = CMP + O-sn-
  phosphatidyl-L-serine.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; POSSIBLE INTERACTION WITH THE
  INNER MEMBRANE (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CDP-ALCOHOL PHOSPHATIDYLTRANSFERASE
  CLASS-II FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL: U32726; AAC22084.1; -.

```

TIGR; H10425; -  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; PLDc; 2.  
 DR SMART; SM00155; PLDc; 2.  
 DR PROSITE; PS50035; PLD; 1.  
 KW Transferase; Phospholipid biosynthesis; Membrane; Repeat;  
 KW Complete proteome.  
 FT DOMAIN 134 160 PLD PHOSPHODIESTERASE 1.  
 FT PIR; S06993; 356 383 PLD PHOSPHODIESTERASE 2.  
 SQ SEQUENCE 455 AA; 53080 MW; 255CA66C1C45B9EB CRC64;  
 Query Match 6.2%; Score 169.5; DB 1; Length 455;  
 Best Local Similarity 21.1%; Pred. No. 8.7e-05;  
 Matches 98; Conservative 73; Mismatches 188; Indels 105; Gaps 17;  
 QY 92 LIESAHSIDLOYIYIRNDISGRLLFNLYLAERG--VRVLLDDNNTGLDILLAL 149  
 DB 42 LIRNAKRRIVYALYKQKDEAGQELDEIYRKQENPHLDVAVLLDM--RAQRNLGAE 99  
 QY 150 DSHPIEVRLEPPEVLRKW---RALGYITDFR-----LNRR-----MHNKFTADR 194  
 DB 100 KSATNAD-----WYCEOROTYOLPDDPNMFEVPIINTREYFGVLYHKGFEVDD- 147  
 QY 195 ATILGRNIGDEYKVEDETFADLDILATSGVGEVSHDEPRYASHS-----HN 246  
 DB 148 TYLYSGASINNYLHOFEXYRDRYOKITHAELADSMVNFINDYLLDSAVYPLDVTNR 207  
 QY 247 ATRIRSGNIGKGLQALGYNDTSRHALLRYREVEOSPPLYOKIOTGRIDMOSVOTRLIS 306  
 DB 208 RTKEIR-GNIRAYRKDLQNGEYSLSKAVKLPNLVSLSPFLGASGN----- 254  
 QY 307 DDPAGLDRDRKRPRIAGRLQDALQOPEKSVYLVSPFYPTSGDALAKLVODGIDVT 366  
 DB 255 -----ELNOVIEDLFLOVKKLYICTPFNFPRPTLOHKIATLLENGKRV 300  
 QY 367 LTNSLATDVAHVHSGYVYKRPPLKAGI--KLYELOPHNAVPARKDGLGSSVTSL-- 422  
 DB 301 IVGKRVANDF-----YIPPEQPFKAGALPYLIESNLRFCEKETOTESQOLVYRLMR 354  
 QY 423 -----HAKTFIVDGKRIFTIGSFNDPRSARLNTMGVIVIESPKIAEQMERTLADTSPEY 476  
 DB 355 DGDNTYHKGWVDDRYILLTGNNLNPRAMRLDAENGILLIYPO-----QQLLAQVEK- 408  
 QY 477 AYRVLDNRNRQWMDPATR-----KTYNPEPAKIMKRIA 512  
 DB 409 -----QNOIRQHTKYLKHYTELELNQYP-EPVOKILMKFA 443  
 RESULT 15  
 YMAP\_STRMU  
 ID YMAP\_STRMU STANDARD: PRT: 373 AA.  
 AC P34001;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein in WAPA 3' region (Fragment).  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 GN NCBI\_TaxID=1309;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=89343633; PubMed=2761387;  
 RA Ferretti J.J., Russell R.R.B., Dao M.L.;  
 RT "Sequence analysis of the wall-associated protein precursor of  
 RT Streptococcus mutans antigen A";  
 RL Mol. Microbiol. 3:469-478(1989).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (potential).  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN  
 CC SYNTHASE SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement and its statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 DR EMBL; M37842; AAA8609.1; -  
 DR PIR; S06993; 506993.  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; PLDc; 1.  
 DR SMART; SM00155; PLDc; 1.  
 DR PROSITE; PS50035; PLD; 1.  
 KW Hypothetical protein; Transferase; Transmembrane.  
 KW TRANSMEM 18 38 POTENTIAL.  
 FT TRANSMEM 124 144 POTENTIAL.  
 FT DOMAIN 197 224 POTENTIAL.  
 FT NON\_TER 373 373 PLD PHOSPHODIESTERASE.  
 SQ SEQUENCE 373 AA; 43410 MW; 220B61BEDB86A84 CRC64;  
 Query Match 5.6%; Score 151; DB 1; Length 373;  
 Best Local Similarity 21.1%; Pred. No. 0.0014;  
 Matches 68; Conservative 49; Mismatches 117; Indels 88; Gaps 12;  
 QY 96 AHSIDLOYIYIRNDISGRLLFNLYLAERGVRVRLIDNNTGLDDL-LALDSHPN 154  
 DB 121 AKKFTLEFEYIIAEGIMMGEILSLEOKYQEGVEVRYWMD-----GMELSTLSFDYAKR 175  
 QY 155 IE-----VLEFNPFLRKWRALGYITDF--PRLNRMHNKSTADNRATILGGRNIGEV 207  
 DB 176 LEKIGIKAKVESP-----ITPVSTYNNYRDHKKIIVIDKKAFFNGINLADRY 224  
 QY 208 F-KVGEDTFVADLDILATSGVGEVSHDFDRY-----ASHANATRIIRSGNIG 257  
 DB 225 INQIERFGWKDFTAVMLBEGVASFTLMFLQMWSTNNQYERAPYLTONFHFIVANGYV- 283  
 QY 258 KGLQALGYNDEFSRHALLRYRETVEOSPPLYOKIOTGRIDMOSVOTRLISDDPAKGLDR 317  
 DB 284 -----IPYSD-----SPL-----DHEKRYGENYI----- 302  
 QY 318 RKPPIAGRLQDALQOPEKSVYLVSPFYPTSGDALAKLVODGIDVYVLTNSLATDVA 377  
 DB 303 -----DILNQARDVYITMPTLILIDSEMAHALQFAERGVDKTIIMPCT--PDKK 350  
 QY 378 AVHSGYVYKRPPLKAGIKLYE 399  
 DB 351 VPFALAKRYFPALDAGVYKE 372  
 RESULT 16  
 PLD2\_RAT  
 ID PLD2\_RAT STANDARD: PRT: 933 AA.  
 AC P70498; O08768;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D2 (EC 3.1.4.4) (PLD 2) (choleine phosphatase 2)  
 DE (Phosphatidylcholine-hydrolyzing phospholipase D2) (PLD2) (rPLD2).  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=98194218; PubMed=9533024;  
 RA Nakashima S., Matsuda Y., Akao Y., Yoshimura S.-I., Sakai H.,  
 RA Hayakawa K., Andoh M., Nozawa Y.;  
 RT "Molecular cloning and chromosome mapping of rat phospholipase D  
 RT genes, Pld1a and Pld1b and Pld2";  
 RL Cytogenet. Cell Genet. 79:109-113(1997).  
 RN [2]  
 RN SEQUENCE FROM N.A., AND CHARACTERIZATION.

CC TISSUE-Brain;  
 CC RX MEDLINE-97269050; PubMed-9111050;  
 RA Kodaki T., Yamashita S.;  
 RT "Cloning, expression, and characterization of a novel phospholipase D  
 RT complementary DNA from rat brain.";  
 RL J. Biol. Chem. 272:11408-11413(1997).  
 RN [3]  
 RN SEQUENCE OF 445-535 FROM N.A.  
 RC TISSUE-Gliol cell;  
 RC RX MEDLINE-96354814; PubMed-8753790;  
 RA Yoshimura S.-I., Nakashima S., Ohguchi K., Sakai H., Shinoda J.,  
 RA Sakai N., Nozawa Y.;  
 RT "Differential mRNA expression of phospholipase D (PLD) isozymes  
 RT during cAMP-induced differentiation in C6 glioma cells.";  
 RL Biochem. Biophys. Res. Commun. 225:494-499(1996).  
 CC -1 FUNCTION: MAY HAVE A ROLE IN SIGNAL-INDUCED CYTOSKELETAL  
 CC REGULATION AND/OR ENDOCYTOSIS (BY SIMILARITY).  
 CC -1 CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -1 ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-  
 CC BISPHOSPHATE, AND PHOSPHATIDYLETHANOLAMINE. INHIBITED BY  
 CC PHOSPHATIDYLSELINE AND BY OLEATE. IS NOT RESPONSIVE TO ADP-  
 CC RIBOSYLATION FACTOR-1 (ARF-1), NOR TO GTP-BINDING PROTEINS: RHO A.  
 CC -1 SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
 CC -1 TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG, HEART, KIDNEY,  
 CC STOMACH, SMALL INTESTINE, COLON, AND TESTIS, AND AT A MUCH LOWER  
 CC LEVELS IN THYMUS, LIVER AND MUSCLE.  
 CC -1 SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.  
 CC -1 SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1 SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC -1 SIMILARITY: CONTAINS 1 PX DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: AB003172; BAA24078.1; -  
 DR EMBL: D88672; BAA19882.1; -  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR00136; PLD.  
 DR InterPro: IPR001683; PX.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00614; PLDC; 2.  
 DR Pfam: PF00787; PX; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00135; PLDC; 2.  
 DR SMART: SM00312; PX; 1.  
 DR PROSITE: PS50003; PH DOMAIN; FALSE\_NEG.  
 DR PROSITE: PS50035; PLD; 2.  
 DR PROSITE: PS50195; PX; 1.  
 DR Hydrolase; Lipid degradation; Membrane; Repeat.  
 KM DOMAIN 65 195 PX.  
 FT DOMAIN 203 311 PH.  
 FT DOMAIN 437 464 PLD PHOSPHODIESTERASE 1.  
 FT DOMAIN 751 778 PLD PHOSPHODIESTERASE 2.  
 FT DOMAIN 441 788 CATALYTIC.  
 FT CONFLICT 26 26 V -> E (IN REF. 2).  
 FT CONFLICT 125 125 N -> P (IN REF. 2).  
 FT CONFLICT 599 599 G -> A (IN REF. 2).  
 FT CONFLICT 792 792 K -> E (IN REF. 2).  
 FT CONFLICT 817 818 GR -> KH (IN REF. 2).  
 FT CONFLICT 919 924 HMGAKR -> PUGSKE (IN REF. 2).  
 SQ SEQUENCE 933 AA; 106036 MW; D430843B4D541EBA CRC64;

Query Match 5.1%; Score 139.5; DB 1; Length 933;  
 Best Local Similarity 18.3%; Pred. No. 0.035;  
 Matches 115; Conservative 92; Mismatches 210; Indels 211; Gaps 27;  
 30 LINCSSS-----WLPPEERTESRHNTSKPVRDLNLIQIR---HTPHTNGLSDIYLL 79

Db 289 LILKCSYQOARWVGGEITELAQG-----PGR--DFQLQHDSYAPRPTLARWVY 339  
 QY 80 NQDHEFAARALIEAESHLDLYI-----WRNDISGRLLFNVLVAE 125  
 Db 340 NGAGYFAAADAILRAREIFITDWWLSPEIYLKRAHSDWRDLI-----MLKRAE 392  
 QY 126 RGVRRVRLDDNNTRGLDULLALDS-----HPNIEVRLFPFVLRKRRALGIL 175  
 Db 393 EGVRSILLFKE-----VELAIGNSYGRKRLMLLHPINKV--MRPDLVTL 439  
 QY 176 DEPRRLRRHNRKSFADNRATILGGRNIG----DEYFK--VGEDTVPADLLATGS- 226  
 Db 440 -----AHHEKLLVDAQVAFLGDLAYGRMDVQYRUTLDGDSSESADSGTPPGSD 492  
 QY 227 --VGVESHSD-----FDRWASHASHNATRIIRSGNIGKGLALGY 265  
 Db 493 PATPDLSSHHEFWLCKDYSNLTCKDWOLDREFEDFIDRETTPRMPWPDGVVHGVAA 552  
 QY 266 NDETSRHALRRETFEGSPLYOKIOTGRIDMSVOT-----RLISDDPAKGLD 314  
 Db 553 RD-LARHFTORNFETTKIARKKIPQYPLLRKSASTANHLFPITIGGCATYQVLRSD 611  
 QY 315 RDRRKPLAGRLQDA-----LKQPKSVYLVSPEYVPTSGDALAKLVODGID--- 363  
 Db 612 RMS-----AGTLESSILNAYLTIRESCHFVLENGFISCSGGRVLRKVGDEIVRL 666  
 QY 364 -----VYVLT-----NSLQATDVAANVSGY----- 383  
 Db 667 KAHEDGOCERVVYLLPLRPFEGEDISTGGNSIQ-----ALHFTYTLGREGYSILRLK 722  
 QY 384 ----VYRRPPLKAGIKELQPNHAVPATKDKGLGSSVTSIAHAKTIVDGRIPTGSE 439  
 Db 723 AAMGTMRWRYMSICGRTGELGHP-----SELIYHSLMLADRTVITGSA 772  
 QY 440 NLDPS--ARLNTMGVTVIESPKIAEOMERTLADISPEYAVRYTLDRHNR----- 487  
 Db 773 NINDRSLCKRSELAIIKID-----TEMEPSIMDGVEYAGRFALSLRCRCSVILGANT 828  
 QY 488 ---LQWHDPAITRKTYNPEAKLWKRIA 512  
 Db 829 WPDLLRDVCDPFF-----QLMDETA 850

RESULT 17  
 PEEL ARATH  
 ID PEEL ARATH STANDARD; PRT; 762 AA.  
 AC Q9C888;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Phospholipase D epsilon (EC 3.1.4.4) (AtPLDepsilon) (PLD epsilon)  
 DE (Pldalpha3).  
 GN PLDPSILON OR ATIG55180 OR F7A10.25.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_Taxid=3702;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE-21016719; PubMed-11130712;  
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,  
 RA White O., Alonso J., Ataifi H., Araujo R., Bowman C.L., Brooks S.Y.,  
 RA Bueller E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
 RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,  
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
 RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,  
 RA Sano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,  
 RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,  
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;  
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:816-820(2000).  
 CC -1- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal  
 CC phosphodiesteric bond.  
 CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -1- COFACTOR: Calcium (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By  
 CC similarity).  
 CC -1- DOMAIN: C2 domain is a calcium-binding fold, and the binding  
 CC promotes the protein association with membranes.  
 CC -1- SIMILARITY: Belongs to the phospholipase D family. C2-PLD  
 CC subfamily.  
 CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, AC027034; AAG51567.1; -;  
 DR InterPro: IPR000008; C2.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDc. 2.  
 DR SMART: SMO0239; C2; 1.  
 DR SMART: SMO0155; PLDc. 2.  
 DR PROSITE: PS50004; C2\_DOMAIN\_2; FALSE\_NEG.  
 DR PROSITE: PS50035; PLD; 1.  
 KW Hydrolyase; Lipid degradation; Calcium; Repeat; Multigene family.  
 FT DOMAIN 13 109 C2 DOMAIN.  
 FT 301 339 PLD PHOSPHODIESTERASE 1.  
 FT DOMAIN 610 637 PLD PHOSPHODIESTERASE 2.  
 FT ACT\_SITE 306 306 POTENTIAL.  
 FT ACT\_SITE 308 308 POTENTIAL.  
 FT ACT\_SITE 313 313 POTENTIAL.  
 FT ACT\_SITE 615 615 POTENTIAL.  
 FT ACT\_SITE 617 617 POTENTIAL.  
 FT ACT\_SITE 622 622 POTENTIAL.  
 SO SEQUENCE 762 AA; 86769 MW; 5E9DA51B463994CC CRC64;  
 Query Match 5.1%; Score 139; DB 1; Length 762;  
 Best Local Similarity 19.4%; Pred. No. 0.028;  
 Matches 129; Conservative 104; Mismatches 203; Indels 228; Gaps 32;  
 OY 20 TRSLISLCLLSCS-----SWLPLEERTESRHPNTSKPVRLNIDLIORTPTHTNGLSID 75  
 DB 121 TKRNKLKLCMLWFRPAYLEPGWCRALEAFSGFGIRNASFPQSRVLYLDYDAHHKATFD 180  
 OY 76 IYLLNDPHEAFAR-----AALIESAEH-----SIDLOYIWRNDIS-----GRL 116  
 DB 181 PRVDVP-----FNARMLMEDYKALIESARHLVYIAGWALNPLVLVDNETETPAHVGTV 237  
 OY 117 FNLVYIAERGVRLLDNDNRGLDLDLLALDSHPNIEVRLFNPFLRK--WRALGYL 174  
 DB 238 GELLKRKEEGVAVRVMANDET-----SLPMIK-----NKGVNRTVEGALAYF 282  
 OY 175 TD-----FPLNRK-----HNKSTADNRAT-----ILGRNIGDEYKYV 210  
 DB 283 RNTNVVYCLRLKHLKPLTAFAHOKITLDRVNSSTKEREIMSGFGFLDCGRYDT 342  
 OY 211 GEDYFAUL-----DLANG-----SVVG-----EYSHDFDYMASH 242

DB 343 EHSLEFRLTEADEFYQTSVAGAKLSRGCPREPMHDCVSVYGAAMDVLKNFEORMTQ 402  
 OY 243 SAHNATRIIRSGNIGKGLQALGYNDETSRHALLRYRETVQ-----SPLYOKIQTG 293  
 DB 403 C--NPSVLVNTSGINLVNLGPTIEENNNKNVQVLRSDIHISATEMPRGLPVEKSHDG 460  
 OY 294 RI-----DMQSVQRLISD-----DEAKGLDRDRKPRPIAGRLQ 327  
 DB 461 YVAAIRKAEFRFYIENQYFMSGDHESKNDKICGCTNLIPVE-----IALKIA 510  
 OY 328 DALKOPK-SYLVSPRYV--PTKSGT-----DALALVQ----- 359  
 DB 511 AKIRAREFFAVYIYIPWPEGESETEVETLHMTRETSMMYQIGRAIWEVGDKSPR 570  
 OY 360 DQIDVTYVTLNSLOARD--VAAVHSGYVYKRPDLKAGIKLVELOPNHAVPATKQGLTGS 417  
 DB 571 DLYNFCLANREKRDGEFAVSSPH-----QKTHVWNAQRNRF-- 610  
 OY 418 SVTSLHAFTFYDCKRIFGSEFNLDPRS--ARLNTMGVIESPKIAQOMERTLADTSPE 475  
 DB 611 -MVVYHSLMTIVDOTYILIGSANINQMSMDCRDEIAIGC-----YQTNNTNEI 661  
 OY 476 YAYRTLDNRHRLQWHDATRK-----TYNPEPA-----KLKRRIA-KILS 517  
 DB 662 QAYRLSL-----WYHTGGKITADLSSESPSELECVRGLRTIGQMWETIYSGDKYVD 714  
 OY 518 LKPI 521  
 DB 715 MGI 718  
 RESULT 18  
 PLD2\_HUMAN STANDARD; PRT; 933 AA.  
 ID 014939; 043540; 043579; 043580; 096BY3;  
 AC 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D2 (EC 3.1.4.4) (PLD 2) (Choline phosphatase 2)  
 GN (Phosphatidylcholine-hydrolyzing phospholipase D2) (PLD1C) (hPLD2).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS PLD2A; PLD2B AND PLD2C), AND  
 RP CHARACTERIZATION.  
 RC TISSUE=Brain;  
 RX MEDLINE=98437320; PubMed=9761774;  
 RA Steed P.M., Clark K.L., Boyar W.C., Lasala D.J.;  
 RT "Characterization of human PLD2 and the analysis of PLD isoform splice  
 RT variants";  
 RL FASEB J. 12:1309-1317(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM PLD2A), AND CHARACTERIZATION.  
 RC TISSUE=B-cell;  
 RX MEDLINE=98250727; PubMed=9582313;  
 RA Lopez I., Arnold R.S., Lambeth J.D.;  
 RT "Cloning and initial characterization of a human phospholipase D2  
 RT (hPLD2). ADP-ribosylation factor regulates hPLD2";  
 RL J. Biol. Chem. 273:12846-12852(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS PLD2A AND PLD2B).  
 RA Seab K.M., Clark J.M., Byrd P.J., Armstrong S.J., Wakelam M.J.O.;  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM PLD2A).  
 RC TISSUE=uterus;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY HAVE A ROLE IN SIGNAL-INDUCED CYTOSKELETAL  
 CC REGULATION AND/OR ENDOCYTOSIS (BY SIMILARITY).



```

CC -1 CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatide.
CC -1 ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-
CC BISPHOSPHATE AND ACTIVATED BY THE ADP-RIBOSYLATION FACTOR-1 (ARF-
CC 1).
CC -1 SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -1 ALTERNATIVE PRODUCTS: 3 ISOFORMS; PLD2A (SHOWN HERE), PLD2B AND
CC PLD2C; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1 TISSUE SPECIFICITY: UBIQUITOUS.
CC -1 SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC -1 SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1 SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -1 SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF033850; AAD04197.1; -
DR EMBL: AF033483; AAC24498.1; -
DR EMBL: AF038440; AAB96655.1; -
DR EMBL: AF038441; AAB96656.1; -
DR EMBL: BC015033; AAH15033.1; -
DR Genbank: HGNC:9068; PLD2.
DR MIM: 602384; -
DR InterPro: IPR001849; PH.
DR InterPro: IPR001736; PLD.
DR InterPro: IPR001683; PX.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00614; PLDc; 2.
DR Pfam: PF00787; PX; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00155; PLDc; 2.
DR SMART: SM00312; PX; 1.
DR PROSITE: PS50003; PH_DOMAIN; FALSE_NEG.
DR PROSITE: PS50035; PLD; 2.
DR PROSITE: PS50195; PX; 1.
KW Hydrolase; Lipid degradation; Membrane; Alternative splicing; Repeat.
FT DOMAIN 65 195
FT DOMAIN 203 311
FT DOMAIN 437 464
FT DOMAIN 751 778
FT DOMAIN 441 788
FT VARSPPLIC 337 933
FT VARSPPLIC 810 820
FT CONFLICT 16 16
FT CONFLICT 165 165
FT CONFLICT 172 172
FT CONFLICT 212 212
FT CONFLICT 268 268
FT CONFLICT 325 325
FT CONFLICT 445 445
FT CONFLICT 510 510
FT CONFLICT 566 566
FT CONFLICT 577 577
FT CONFLICT 632 632
FT CONFLICT 784 784
FT CONFLICT 801 801
FT CONFLICT 821 821
SQ SEQUENCE 933 AA; 105986 MW; F25F6B73B45F57ED CAC64;
Query Match 5.1%; Score 137.5; DB 1; Length 933;
Best Local Similarity 18.2%; Pred. No. 0.048;
Matches 118; Conservative 91; Mismatches 186; Indels 255; Gaps 31;
QY 30 LSCSS-----WLPPLERTSRHFNKSPYRLDNIQI-RH---TPTNLSIYLL 79
DB 289 LTKCSYQARWMAOEITELAG-----PGR--DFQLHRHDSYAPPRGTLARWV 339

```

```

QY 80 NDEHEFAARAAALIESAEHSLDQYI-----WRNDISGRLLFNLYIAE 125
DB 340 NGGVFAAAYADALIRNOEELFTIDWLSPEVYLAKRAHSDWLDI-----MLRKAE 392
QY 126 RGVYRLLDDNNTRGDLDLLALDS-----HPNIEVRLNPEVLRKRALGYLT 175
DB 393 EGVRSYLLTFKE-----VELALGINSYKRALMLLPNIKV-MRRPDQTLW-----439
QY 176 DFLRLRRMHNKSFADNRATILG-----RNIGD-----205
DB 440 -----AHHEKLLVYDQVAVFLGLDLAYGRMDLHYRLTLDGSSASAOPPPRPD 492
QY 206 -----EYKGVGD-----TVFADLLIAGSVY 228
DB 493 SPATPDLNHNQFELWKDYSNLITKDWQDLRFEDFIDRETPRMPWVGYNVGLA 552
QY 229 GEVSHDPDRYKASHASHNATRIIRSGNIGKGLDALGYNDSTSHALLRREYQSPLYO 288
DB 553 RDLARHFQIRW-----NFKTKTKA-----KYKTPPYLLPKSTSTANQLPF--594
QY 289 KIQGRIDMQSVOTRLISDPAKGLDRDRKRPRIAGRLDA-----LKQPEKSVYLV 340
DB 595 TLPGGQC--TVQV-----LRSVDRWS-----AGTLESLINAVLHTYRESQHFLYTE 640
QY 341 SPYFVPTKSGTDAIAKLVDQDID-----VTYLT-----NSLQ 372
DB 641 NQFISCSGRVYTLNKGVDIVRIILKAHQGCYRYVYLLPLPPEGDISGCGNSIQ 700
QY 373 ATDVAAVHSGY-----VKTRKPLKAGIKLYELOPNHAVPATKDKG 413
DB 701 -----AIIHFYRTLGRGEYSILHRLKAMGTAMRDYISIGLTHGELGHPV-----749
QY 414 LTGSSVTSIAKTFIVDKRIFGFSFNLDPRS--ALNTNEMGVYIE-----SPRI--ABQ 464
DB 750 ----SELIYHSKVLADRDVYIIGSANINDRSLGKRDSLAVALIEDTETEPILMNGAEY 806
QY 465 MERTLADTSEPVAYRYTLDRHNR--LQWMDPATRKTYPNPEAKLWKRIA 512
DB 807 QACRFALSLRKHCFGVILGANTRPDLDRPICDDPF-----QLWQDMA 850
RESULT 19
PSS_ECOLI STANDARD; PRT; 451 AA.
AC P23830; P78100; P78256;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE CDP-diacylglycerol--serine O-phosphatidyltransferase (EC 2.7.8.8)
DE (Phosphatidylserine synthase).
OS PSSA OR PSS OR B2585.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-12.
RX MEDLINE=91161632; PubMed=2002065;
RA Dechavigny A., Heacock P.N., Dowhan W.;
RT "Sequence and inactivation of the pss gene of Escherichia coli.
RT Phosphatidylethanolamine may not be essential for cell viability.";
RL J. Biol. Chem. 266:5323-5332(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]

```

SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RC MEDLINE-97349980; PubMed-9205837;  
 RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
 RA Oshima T., Oyama S., Saito N., Samped G., Satoh Y., Sivasubram S.,  
 RA Tadami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
 RA Yamagata S., Horuchi T.;  
 RA "Construction of a contiguous 874-kb sequence of the Escherichia coli  
 RT - K12 genome corresponding to 50,0-68.8 min on the linkage map and  
 RT analysis of its sequence features.";  
 RL DNA Res. 4:91-113(1997).  
 RN [4]  
 RP REVIEW.  
 RA MEDLINE-92356873; PubMed-1323044;  
 RA Dowhan W.;  
 RT "Phosphatidylserine synthase from Escherichia coli.";  
 RL Meth. Enzymol. 209:287-298(1992).  
 CC -1- CATALYTIC ACTIVITY: CDP-diacylglycerol + L-serine -> CMP + O-sn-  
 CC phosphatidyl-L-serine.  
 CC -1- SUBUNIT: MULTIMERIC.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; POSSIBLE INTERACTION WITH THE  
 CC INNER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE CDP-ALCOHOL PHOSPHATIDYLTRANSFERASE  
 CC CLASS-II FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M58699; AAA97504.1; -;  
 DR EMBL: AE000345; AAC75638.1; ALT\_INIT.  
 DR EMBL: D90886; BAA16470.1; ALT\_INIT.  
 DR EMBL: D90887; BAA16473.1; ALT\_INIT.  
 DR PIR: JH0368; JH0368.  
 DR Ecogene; EG10781; pssa.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 DR PROSITE: PS50035; PLD; 2.  
 DR Transferrase; Phospholipid biosynthesis; Membrane; Repeat;  
 KW Complete proteome.  
 FT DOMAIN 133 159 PLD PHOSPHODIESTERASE 1.  
 FT DOMAIN 352 379 PLD PHOSPHODIESTERASE 2.  
 FT DOMAIN 4 9 LYS-RICH (BASIC).  
 FT DOMAIN 119 154 HYDROPHOBIC.  
 FT DOMAIN 239 284 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 430 449 ARG/LYS-RICH (BASIC).  
 FT CONFLICT 32 32 A -> R (IN REF. 1).  
 FT CONFLICT 78 78 R -> DD (IN REF. 1).  
 FT CONFLICT 165 167 K YR -> NIA (IN REF. 1).  
 FT CONFLICT 287 288 LL -> FV (IN REF. 1).  
 FT CONFLICT 309 309 P -> S (IN REF. 1).  
 SQ SEQUENCE 451 AA; 52801 MW; 9E9A2A5C4B4C814F CRC64;

Query Match 4.8%; Score 130; DB 1; Length 451;  
 Best Local Similarity 21.0%; Pred. No. 0.058;  
 Matches 98; Conservative 68; Mismatches 169; Indels 132; Gaps 23;

65 RHTPTNGSLDYLNDPEAFRA---RAALIE---SAEHSLDLYYIMRNDISGRILE 117  
 Db 9 KHQHLAQLPKISQSDVDVDFYAPADRETIIEKIASKKQICVALYLEDDGKGKIL 68  
 118 NLVYLAER--GVRVRLLD-----DNNTRGLDILLALDSHPNIEVRLFN-P 162  
 Db 69 NALYEAKKQRPDLVAVLVDMHRAQGRIGAAASNTNA-DWCRMAQENPGVDVYVYGV 127

163 EVLRKRALGYLTDPRLNRMRHNSFTADNRATILGNNIDEFYKGEDEVFADLDIL 222  
 Db 128 INTRE--ALGLV-----HFKEFIID--SVLYSGASLNDVYLH----- 161  
 223 AGSGVAGEVSHD---FDRY-----WASHSHMNTFRIIRSNIGKGLQALGY 265  
 Db 162 -----QHDKYRDYRHLIRNRKMSDIMEFWYQNTMN-----GRGVNR-- 200  
 266 NDETSRHALLRREVEQSP-LYQKIQTGRIDMQSQRTLISDDPAKGLDRDRRK-PP1A 323  
 Db 201 -DDVNR-----PKSPLEKNIDIRLEPROELRDAAYFGQD-----ADNQGLSVPLV 244  
 324 GRLQDALKO-----PEKSVLYSPYFVPTKSGTDALAKLVODGIDVYVINSLOAT 374  
 Db 245 GKGSSLLNKRTIFHLMPCAEOKLTICTPYFNLPAILVRNIOLLREGKKEVIEVDKTN 304  
 375 DVAANVHSGVYKRRP--LLKAGIKLEYLOPNNAVPATKXGLTGSYSVL-----HA 424  
 Db 305 DE-----YIPDEPKRTIIGALPYLEINLRRFLSRLOQYVNTDOLVRLMKDDNTYHL 358  
 425 KTFIYDGRKRIFGSEFLDPRSRALNTMGVIESP--KIAEQMERTL 469  
 Db 359 KGMWVDDKMLITGNLNNLRAMRDLLENAILIHDPOLELAPQREKL 405

RESULT 20  
 SP14\_YEAST  
 ID SP14\_YEAST STANDARD; PRT; 1380 AA.  
 AC P36126;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D1 (EC 3.1.4.4) (PLD 1) (Choline phosphatase  
 DE 1) (Phosphatidylcholine-hydrolyzing phospholipase D1) (Meiosis-  
 DE specific sporulation protein SP014).  
 GN SP014 OR PLD1 OR YKR031C.  
 OS Saccharomyces cerevisiae (baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Uresterazu L.A., Janniaux J.-C.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP MEDLINE-92258685; PubMed-1582554;  
 RA Honigberg S.M., Conicella C., Esposito R.E.;  
 RT "Commitment to meiosis in Saccharomyces cerevisiae: Involvement of  
 RT the SP014 gene.";  
 RL Genetics 130:703-716(1992).  
 RN [3]  
 RP CHARACTERIZATION AS A PLD.  
 RX MEDLINE-96109223; PubMed-8618862;  
 RA Rose K., Rudge S.A., Frohman M.A., Morris A.J., Engbrecht J.;  
 RT "Phospholipase D signaling is essential for meiosis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:12151-12155(1995).  
 RN [4]  
 RP CHARACTERIZATION AS A PLD.  
 RX MEDLINE-96161959; PubMed-8576189;  
 RA Maksman M., Eli Y., Lisovitch M., Gerst J.E.;  
 RT "Identification and characterization of a gene encoding phospholipase  
 RT D activity in yeast.";  
 RL J. Biol. Chem. 271:2361-2364(1996).  
 CC -1- FUNCTION: REQUIRED FOR MEIOSIS AND SPORE FORMATION. SEEMS TO BE  
 CC INVOLVED IN THE COORDINATE INDUCTION OF LATE MEIOTIC EVENTS. PLD  
 CC ACTIVITY IS INDUCED UNDER SPORULATION CONDITIONS AND SEEMS TO BE  
 CC NECESSARY TO COMPLETE THE MEIOTIC CYCLE, BUT NOT FOR VEGETATIVE  
 CC CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -1- ENZYME REGULATION: ACTIVITY IS DEPENDENT OF PHOSPHATIDYLINOSITOL  
 CC 4,5-BISPHOSPHATE. INHIBITED BY MAGNESIUM.

Query Match	4.7%	Score	128.5	DB	1	Length	1380
Best Local Similarity	17.9%	Pred. No.	0.38				
Matches	98	Conservative	97	Mismatches	199	Indels	153
						Gaps	25

```

OY 15 ISLKTSTSLISLILLLLOCSWSSWLPPEERFESRHFNTSKVRLDNIQIIOHTPTNLS 74
Db 328 LTLNSEKKTITICKSESLKQMMSSITKMTSTPW--SKFRNGSFAPVR---TNSFC 381
OY 75 DYLNDPHEAFARAAALIESAE---HS-LDIQYIWRNDISGRLLF---NLYUAAE 125
Db 382 K-FLVGDHDPWMSLEALMAKVYIYIDMWMLSPELY-RRPVKNGQGFRIIDMLSCAE 439
OY 126 RGVAVRLLDON--NTRLDLULL--ALDSHPNIEVNLFPVLRKKRRLGYLTFDFRL 180
Db 440 KGIKIFVIYIYVNGVIGTDSLMTKHSMLNHPNINI--TRSP---NOMLONTYFW---- 490
OY 181 NRRMHNKSTFADNRPATILGSRNIGDEYKRVGEDVFADLDILATGSVGE-----VS 232
Db 491 --AHHEKRVVIDEFAFICGTDLICGRVDTEFENHVRDDAESLLDQNPFGKQVSNARIADF 548
OY 233 HDDEPTWAS-----HSAHNAITRISGNIGCGLQALGYNDETSRHALLYR-- 278
Db 549 HDLDKPFESMYDRKVIYRPMRHVDQMMI-----LGEPARDL-----ARHFYWRMYL 595
OY 279 -----ETVEQSPLOYKIOIGRIDMOSVQIRLISDDPAKGLDR 317
Db 596 LRAKRPSHLPTLTPRPDSDLTAEELKSLRPFELR-----EKSTCEIQLISAGNWSLGLKE 651
OY 318 RKRPTAGLQALQKPEKSVTVLSEFYVPYTK--SSTDALAKLIVODGIDVYVLTN----- 369
Db 652 TEGSIONAYLKLIEOSEHFYIYIENOFFITSTYVMNCTVLRNIGDALVDRIVYKANEKKRPW 711
OY 370 -----SLOATDVAAY-----HSQYVYRKPDLKAG--IKLY 398
Db 712 KAFILIPMLPGFDSVPDVAEASSLRLINQFOYQOSTSRBEHSTFSKLKLINIDPAQYIOQF 771
OY 399 ELQ-----PNHAAVPATKQGLTGGSSVTS--LHAKTFIVDGKRIEIGSFNLDPPSARLN 449

```

```

Db      772 SLRKSTAPNRL-----ITBOLYVHAKILLADRCLIGSANINERSQLCN 819
QY      450 TEMGYI 456
       : |
Db      820 RDSEVAI 826

RESULT 21
PLD2_MOUSE STANDARD: PRT: 933 AA.
AC ID PLD2_MOUSE STANDARD: PRT: 933 AA.
P97813:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphatidase D2 (EC 3.1.4.4) (PLD 2) (Choline phosphatase 2)
DE (Phosphatidylcholine-hydrolyzing phospholipase D2) (PLDLC) (mPLD2).
GN PLD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo, and Neonatal brain;
RA MEDLINE=971199446; PubMed=9395408;
RX Colley W.C., Sung T.-C., Roll R., Jenco J.M., Hammond S.M.,
RA Altschuller Y.M., Bar-Sagi D., Morris A.J., Frohman M.A.;
RT "Phospholipase D2, a distinct phospholipase D isoform with novel
RT regulatory properties that provokes cytoskeletal reorganization.";
RL Curr. Biol. 7:191-201(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=N129;
RX MEDLINE=98228223; PubMed=9560313;
RA Redina O.E., Frohman M.A.;
RT "Organization and alternative splicing of the murine phospholipase D2
RT gene."
RL Biochem. J. 331:845-851(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=97493716; PubMed=9307024;
RA Colley W.C., Altschuller Y.M., Sue-Ling C.K., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Branch K.D., Tsifra S.E., Bollag R.J.,
RA Boilag W.B., Frohman M.A.;
RT "Cloning and expression analysis of murine phospholipase DL.";
RL Biochem. J. 326:745-753(1997).
CC -1- FUNCTION: MAY HAVE A ROLE IN SIGNAL-INDUCED CYTOSKELETAL
CC REGULATION AND/OR ENDOCYTOSIS.
CC -1- CANALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphate.
CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-
CC BISPHOSPHATE. IS NOT RESPONSIVE TO ADP-RIBOSYLATION FACTOR-1 (ARF-
CC 1), NOR TO GTP-BINDING PROTEINS: RHO A.
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBQUITOUS. HIGHEST LEVELS IN BRAIN AND LUNG.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS IN THE HIPPOCAMPUS
CC AT THE EARLIEST TIME AT WHICH IT IS DEFINED AS A STRUCTURE AND
CC ALSO IN VENTRICULAR NEURAL CELLS AS WELL AS DIFFERENTIATING
CC NEURONS OUTSIDE OF THE VENTRICULAR REGION. EXPRESSED DURING
CC DEVELOPMENT IN LOWER LEVELS IN MESENCHYMAL CELLS DERIVED FROM THE
CC NEURAL CREST THAT ARE DESTINED TO FORM BONES OF THE SKULL.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U87557; AAC53173.1; .  
 DR EMBL: AF052294; AAC24519.1; .  
 DR EMBL: AF052291; AAC24519.1; JOINED.  
 DR EMBL: AF052293; AAC24519.1; JOINED.  
 DR EMBL: AF052292; AAC24519.1; JOINED.  
 DR MGI: MGI:892877; PLD2.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR001736; PLD.  
 DR InterPro: IPR001683; PX.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00614; PLDc; 2.  
 DR Pfam: PF00787; PX; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00155; PLDc; 2.  
 DR SMART: SM00312; PX; 1.  
 DR PROSITE: PS50003; PH.DOMAIN; FALSE\_NEG.  
 DR PROSITE: PS50035; PLD; 2.  
 DR PROSITE: PS50195; PX; 1.  
 DR Hydrolase; Lipid degradation; Membrane; Alternative splicing; Repeat.  
 DR KW DOMAIN 65 195 PX.  
 FT DOMAIN 203 311 PH.  
 FT DOMAIN 437 464 PLD PHOSPHODIESTERASE 1.  
 FT DOMAIN 751 778 PLD PHOSPHODIESTERASE 2.  
 FT DOMAIN 441 788 CATALYTIC.  
 SQ SEQUENCE 933 AA; 106167 MW; 106167 MD; BADJED0F2EAC9ED CRC64;

Query Match 4.7%; Score 127; DB 1; Length 933;  
 Best Local Similarity 18.6%; Pred. No. 0.27; Indels 178; Gaps 25;  
 Matches 98; Conservative 86; Mismatches 164;

OY 86 FAARALIESAHS-----LDLOYI-----WRNDISGRLEFNLYIAERGVRV 130  
 DB 345 FAADVADALRAQOEIFITDMWLSPEITLKRPAHSDDMRLDI-----MLKRAEGVAV 397  
 OY 131 RLDDNNRGRGDDLLALDS-----HPNIEVRLFPNPLKRWALGLYLDFFPL 180  
 DB 398 SILLKE-----VELALGINSYKRTMLLHPNKKV-MRHPDLVYLW----- 439  
 OY 181 NRRMNRKSTADNRATILGGRNIG-----DEYFK-----VEGDIYVADLILATGS-----VVG 229  
 DB 440 --AHHEKLLVNOVYVAFGLGDLAFGRMDVOYRLDGLDSEPHYLQTPITLCSDEPAATP 497  
 OY 230 EVSHD-----FDRYWASHSAHNATRIIRSGNIGKGLQALGYNDETS 270  
 DB 498 DLSHQFELGKDYSLTKDWVQDLRPFEDFIDRETFRRMWRDVGAVVHGAARD-LA 556  
 OY 271 RIALLRIRTEVE-----GSPLYQKI-----OTGRIDQOSVOTRLISDPKAGLDRDRK 319  
 DB 557 RHFIQRWNETTKYRKATKPYLLPKSTANNLPFMIPIGOCATYVLRSSVDRMS-- 614  
 OY 320 PPIAGRLQDA-----LKQPEKSYLYVSPYFVPTKSGTDALAKLYOGCID----- 363  
 DB 615 ---AGTLESLNATLHTITRESQHFLEYENOFFISCSDSRTVYLNKGVGDIYRILKAHQ 671  
 OY 364 -----VTVLT-----NSLQATDVAAVHSGY-----V 384  
 DB 672 GGCFFVYLLPLLPFGEGDISTGGSNSTO---ALIHFFYRLKGEHSILRLKAAMKT 727  
 OY 365 KYRKLPLAKGKLYELQPHNAVPAATKDKGLGSSVTSLSHAKTFIVDGKRIPTGSENLDR 444  
 DB 728 AWRDYMSTICGLRTHGELGHPI-----SELIIHSMKMLIADDDRTVITIGSANINDR 777  
 OY 445 S--ARLNTNEMGVIESPKAQBOMERTIADTSPEY---ARVYLDGH 485  
 DB 778 SLUGRDSLELATLIRD---TEPEPSLMD-GVEYQAGRPAISLRKH 818

RESULT 22  
 PDL1\_CRIGR

ID PDL1\_CRIGR STANDARD: PRT; 1036 AA.  
 AC 008684;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Phospholipase D1 (EC 3.1.4.4) (PLD 1) (Choline phosphatase 1)  
 DE (Phosphatidylcholine-hydrolyzing phospholipase D1).  
 GN  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;  
 OC Cricetus  
 OX NCBI\_Taxid=10029;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA Klaus J.R., Baldassare J.J., Raben D.M.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -FUNCTION: IMPLICATED AS A CRITICAL STEP IN NUMEROUS CELLULAR  
 CC PATHWAYS, INCLUDING SIGNAL TRANSDUCTION, MEMBRANE TRAFFICKING, AND  
 CC THE REGULATION OF MITOSIS. MAY BE INVOLVED IN THE REGULATION OF  
 CC PERINUCLEAR INTRAVESICULAR MEMBRANE TRAFFIC (By similarity).  
 CC -CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a  
 CC phosphatidate.  
 CC -ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-  
 CC BISPHOSPHATE AND PHOSPHATIDYLINOSITOL 3,4,5-TRISPHOSPHATE,  
 CC ACTIVATED BY THE PHOSPHOKINASE C-ALPHA, BY THE ADP-RIBOSYLATION  
 CC FACTOR-1 (ARF-1), AND TO A LESSER EXTENT BY GTP-BINDING PROTEINS:  
 CC RHO A, RAC-1 AND CDC42 (By similarity).  
 CC -SUBCELLULAR LOCATION: PERINUCLEAR REGIONS; ENDOPLASMIC RETICULUM,  
 CC GOLGI APPARATUS AND LATE ENDOSOMES (By similarity).  
 CC -SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.  
 CC -SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.  
 CC  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

Query Match 4.5%; Score 122; DB 1; Length 1036;  
 Best Local Similarity 19.7%; Pred. No. 0.72; Indels 200; Gaps 31;  
 Matches 117; Conservative 83; Mismatches 193;

OY 60 NILQIRHTPTNGSLDIY-----LNDPHEAFARALIESAHSGLDLOYITWRND- 110  
 DB 444 NIKVNRHDPHVS--SSVILMAHHEKLYIIDQSVAVG-----GIDLATGWRDNE 491

```

QY 111 -----ISGRLLFNVLVAERGVRLLLDDNNTRGIDLLALDS--HPNI--- 155
DB 492 HRLTDGVSKRYTSGLSMGSLLAATMESMESLSLKNHSHKNEPRLKSVDDVDPRLKCV 551
QY 156 -EVRLFNPNVLRKWRALGYLTPPRLNRR-MHNKSFADNRATILGGRNIDEFYKVED 213
DB 552 GKRPFESKFSLL-----YROLHRRHLN-----SDSVSSIDSASNTG----- 587
QY 214 TVFADLDIATSGVYGEVSHDRYMAHSAIN-----ATRIIRS 253
DB 588 -----SIRSVOTGVEL-HGETRWHGKDYCNFVKWVQDLKPAPFIDRYSTPIRW 640
QY 254 GNIGKGLDALGYNDSTSHALLRYRETVQSPLYOKIO-----TGRIIMO----- 298
DB 641 HIGISVLHKAARD-VARHFIQRMNTKIMKPKYRSLSPFLPKSQSTAHRLRYVPGA 699
QY 299 ---SVQTRLLSDPAKGLDRRRKPRPIAGRLDALKOPKSVLYSPVYP-----T 347
DB 700 VPAKVOLLRSADMSAGI---KHHEESIHSAVINVIENSKHYIEMQFISCADKRVFN 757
QY 346 KSGTALAKLY-----QDGDVTVLT---NSIQTADVAVHSGYVXY 386
DB 758 KVG-DALVQRIILKAHREGORVYIYIPRLPGEGDISTGGNALQAI-----MHENT 809
QY 387 RRPRLKAGIKLYELQF-----NHAVPATRKQGLGSSVYTSL---HAKTF 427
DB 810 RTMCGENGSLILQKRLKELGNOMINYSFCGLRTHA-----ELBENLVTELLYVHSKLL 862
QY 428 IYDGKRIFFGSPNLDPRS--ARLNTDMGVIES-----PRIAOME---RTLADTSEPIAY 478
DB 863 IADMTVIIIGSANINDRSMIGKRDSEMAVIYDTEVTPIMDKKEYQACFCAGLRLQCF 922
QY 479 RVTLL-----DRHNRLOWHDPATRKTYPNDEPAKLMKRIIA-----KITSLIP 520
DB 923 RLVLGLSPSEDLO--DPSVDKF-----KEIWSSTAARNATITYKVERCLP 968

RESULT 23
ALAB_ARATH STANDARD: PRT: 1203 AA.
ID ALAB_ARATH
AC Q9SAF5:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Potential phospholipid-transporting ATPase 11 (EC 3.6.3.1)
DE (Aminophospholipid flippase 11).
GN ALA11 OR AT1G13210 OR F3F19.24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Eguu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Lutos J.S., Maici R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Niernan W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
"Sequence and analysis of chromosome 1 of the plant Arabidopsis

```

```

RT thaliana."
RL Nature 408: 816-820(2000).
CC -1- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CARLON TRANSPORT ATPASES FAMILY
CC (E1-E2 ATPASES). SUBFAMILY IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC007357; AAD31074.1;
CC InterPro: IPR001757; ATPase_E1-E2.
CC InterPro: IPR001454; Hlgase/hydriase.
CC Pfam: PF00702; Hydrolase: 1.
CC PRINTS: PR00119; CATAPPAE.
CC PROSITE: PS00154; ATPASE_E1_E2: 1.
CC Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
CC Magnesium; Multigene family.
CC DOMAIN 1 71
CC TRANSMEM 72 93
CC DOMAIN 94 97
CC TRANSMEM 98 120
CC DOMAIN 121 303
CC TRANSMEM 304 325
CC DOMAIN 326 363
CC TRANSMEM 364 381
CC DOMAIN 382 921
CC TRANSMEM 922 941
CC DOMAIN 942 955
CC TRANSMEM 956 975
CC DOMAIN 976 1005
CC TRANSMEM 1006 1028
CC DOMAIN 1029 1041
CC TRANSMEM 1042 1064
CC DOMAIN 1065 1071
CC TRANSMEM 1071 1091
CC DOMAIN 1092 1108
CC TRANSMEM 1109 1133
CC DOMAIN 1134 1203
CC MOD.RES 429 429
CC METAL 866 866
CC METAL 870 870
CC SEQUENCE 1203 AA: 136582 MW: D06ED2FBA046FC6E CRC64;

Query Match 4.1% Score 112; DB 1; Length 1203;
Best Local Similarity 22.9%; Pred. No. 4.7; Mismatches 102; Gaps 20;
Matches 88; Conservative 54; Mismatches 140; Indels 102; Gaps 20;

QY 125 ERGVRVRLDLDNNTRG--DDLALLD--SHPIE-----VRLFPVLRK 167
DB 459 ERSMAR-----SNGSLVGDLDVVDGSPKIKGFNFLDERVYMGWVKKQDAVLOK 513
QY 168 -WALGYL-TDFRLRNRHMKSFTA---DNKATILGGRNIDEFYKVEDTY-PADLDI 221
DB 514 FFLLLVCHVAIETDEAGSVSYEASDPDEAFVAARFEGFESRQNGISFLD- 572
QY 222 LATGVVGEVSHDFDYYMAHSAHNTIRISGN-----IGKGLQAL----- 263
DB 573 LASGKTVERYRLNLVLFNSAKRNSVYRDEGRLILLSSKADNVMPERLAKNGRKE 632
QY 264 -----GYNDETSHALLRYRETVQSPLYOKIOTGRIDMOSVQTRLLSDPAKGLDR 315
DB 633 EKTREHVNERYADAGRLTLILAYREVDE-----EYIEFSKNFNKASVTA 678
QY 316 DRKPRPIAGRLDALKOPKSVLYSPVYPTR--SGT-DALAKLVQDGDVTVLTNSLO 372
DB 679 DRE-----SLIDEITQEMERDILLGATAVEDKLONGVPCIDKLAQAGIKIWLTKDM 733

```

```

OY 373 ADVAAVHSGY-----VKYKPLKAGIK-LYEOPHNHVPATDKG- 413
DB 734 ER---AINIGFACSLROEMKOIITNLETPHIKALEKAGEKAIERASRESVYNOMECK 790
OY 414 --LTGSVTSLSH-AKTEIVDGKRI 434
DB 791 ALTTASSSSASSHFAFLLIDGKSL 814

RESULT 24
PLD_STRAT STANDARD: PRT: 556 AA.
AC 053728:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase D precursor (EC 3.1.4.4) (Choline phosphatase).
OS Streptomyces antibioticus.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1890;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=95151292; PubMed=7765769;
RA Iwasaki Y., Nakano H., Yamane T.;
RT "Phospholipase D from Streptomyces antibioticus: cloning, sequencing,
RT expression, and relationship to other phospholipases."
RL Appl. Microbiol. Biotechnol. 42:290-299(1994).
CC -1 CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatidate.
CC -1 SUBCELLULAR LOCATION: Secreted.
CC -1 SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; D16444; BAA03913.1;
CC Interpro: IPR001736; PLD.
CC Pfam: PF00614; PLDc. 2.
CC SMART: SM00155; PLDc. 2.
CC PROSITE: PS50035; PLD; 2.
CC HYDROLASE; Lipid degradation; Signal; Repeat.
CC SIGNAL
CC CHAIN 1 47
CC FT 48 556 PHOSPHOLIPASE D.
CC FT 210 237 PLD PHOSPHODIESTERASE 1.
CC FT DOMAIN 484 511 PLD PHOSPHODIESTERASE 2.
CC SEQUENCE 556 AA; 58932 MW; B388F0BDC0A596 CRC64;

Query Match 4.1%; Score 111.5; DB 1; Length 556;
Best Local Similarity 20.2%; Pred. No. 1.7;
Matches 76; Conservative 47; Mismatches 120; Indels 133; Gaps 18;

OY 185 HNKSTADNRATITLGRN-IGDEYKVEDIV--FADIDLATGSGVGEVSHDPRYAS 241
DB 215 HSKLLVVGKTAITGSGMKDDYL---DTAHVSDVMALSGPAASAGKYLDTLW-D 269
OY 242 HSAHNAAT-----RIIRSGNICKGLO----- 261
DB 270 WTCRASDPARVWLATNSGASCMPSEDEGASAPETGDPVPIAVGGLGVIGESDPS 329
OY 262 -----ALGYNDETSRHALLRYRETV--EOSPXYKIQIGRIDMOSVQRL 304
DB 330 SGYHDLPTAPDKCYVLLHNTAND---KDYDYNPEENALRSILASARSHVEISQ-- 384
OY 305 ISDDPAKGLDRDRKRPPIAGRLDALKQPEKSVYLVSPYFVPTKSGTALAKLVVDGIDV 364
DB 385 -----DINATCPPL-----PR---YDIRTY-----DTLAGKILAAGYKV 414

```

```

OY 365 TYVNSLQATDVAAVHSGYKRYKPLKAGIKLYELOPHNHPATDKD-----KGLTGS-- 417
DB 415 RIYVSD-PANRGAVSGGYSQ-----IKSLDEISDTLFRRLVALGTDNKKASALQGNQ 468
OY 418 -----SVTSLHAKTFLVQDKRIFGFSNLDPSARLNTKMGVYIESPKTA 462
DB 469 LASFRSSDAKMAKDGKPYVLLHKLIVSDSAFYIGSKNLYPAMLO---DFGYVESPA 525
OY 463 EQMERTLADTSPEYAY 478
DB 526 QQLKTFELD--PEWKY 539

RESULT 25
TYCC_BACBR STANDARD: PRT: 6486 AA.
AC 030409:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrocidine synthetase III (Includes: ATP-dependent asparagine
DE adenylyase (AsnA) (Asparagine activase); ATP-dependent glutamine
DE adenylyase (GlnA) (Glutamine activase); ATP-dependent tyrosine
DE adenylyase (TyrA) (Tyrosine activase); ATP-dependent valine adenylyase
DE (ValA) (Valine activase); ATP-dependent ornithine adenylyase (OrnA)
DE (Ornithine activase); ATP-dependent leucine adenylyase (LeuA) (Leucine
DE activase)).
GN TYCC.
OS Bacillus brevis.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX NCBI_TaxID=1193;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 8185;
RX MEDLINE=98012987; PubMed=9352938;
RA Moez H.D., Marahel M.A.;
RT "The tyrocidine biosynthesis operon of Bacillus brevis: Complete
RT nucleotide sequence and biochemical characterization of functional
RT internal adenylylation domains."
RL J. Bacteriol. 179:6843-6850(1997).
CC -1 FUNCTION: INCORPORATES SIX AMINO ACIDS (FOR TYROCIDINE A, ASN,
CC GLN, TYR, VAL, ORN, AND LEU) IN THEIR L-CONFIGURATION INTO THE
CC PEPTIDE PRODUCT.
CC
CC -1 COPFACTOR: CONTAINS 6 COVALENTLY BOUND PHOSPHOPANETHEINES (BY
CC SIMILARITY).
CC -1 PATHWAY: Cyclic peptide antibiotic tyrocidine biosynthesis.
CC -1 SUBUNIT: LARGE MULTIMERIC COMPLEX OF TYCA, TYCB AND TYCC.
CC -1 DOMAIN: CONSISTS OF SIX MODULES, AND HARBORS A PUTATIVE
CC THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC ADENYLYATION, THIOYLATION, CONDENSATION (NOT FOR THE INITIATION
CC MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC (OPTIONAL).
CC
CC -1 MISCELLANEOUS: TYROCIDINE IS A MIXTURE OF FOUR CYCLIC
CC DECAPEPTIDES, TYROCIDINE A (D-PHE-PRO-PHE-D-PHE-ASN-GLN-TYR-VAL-
CC ORN-LEU), B, C, AND D, IN WHICH PHE, AT POSITIONS 3, 4, AND TYR
CC RESIDUES ARE GRADUALLY REPLACED BY TRP, DEPENDING ON THE RELATIVE
CC CONCENTRATIONS OF THESE AMINO ACIDS IN THE GROWTH MEDIUM.
CC -1 SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
CC
CC -1 SIMILARITY: CONTAINS 6 ACYL CARRIER DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF004835; AAC45930.1;

```

DR HSP: P14687; 1AMU.  
 DR InterPro: IPR000873; AMP-bind.  
 DR InterPro: IPR001242; Condensate.  
 DR InterPro: IPR003880; Pantane\_attach.  
 DR InterPro: IPR000379; Ser-ester\_site.  
 DR InterPro: IPR001031; Thioesterase.  
 DR Pfam: PF00501; AMP-binding; 6.  
 DR Pfam: PF00550; PP-binding; 6.  
 DR Pfam: PF00668; Condensation; 6.  
 DR Pfam: PF00975; Thioesterase; 1.  
 DR PRINTS: PR00154; AMPBINDING.  
 DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 6.  
 DR PROSITE: PS00455; AMP\_BINDING; 6.  
 DR PROSITE: PS50075; ACP\_DOMAIN; 6.  
 DR Ligase: Antibiotic biosynthesis; Phosphopantetheine;  
 KW Multifunctional enzyme; Repeat.  
 FT REPEAT 466 1038 DOMAIN 1 (ASPARAGINE-ACTIVATING).  
 FT REPEAT 1521 2070 DOMAIN 2 (GLUTAMINE-ACTIVATING).  
 FT REPEAT 2536 3113 DOMAIN 3 (TYROSINE-ACTIVATING).  
 FT REPEAT 3590 4149 DOMAIN 4 (VALINE-ACTIVATING).  
 FT REPEAT 4606 5203 DOMAIN 5 (ORNITHINE-ACTIVATING).  
 FT REPEAT 5658 6245 DOMAIN 6 (LEUCINE-ACTIVATING).  
 FT DOMAIN 970 1037 ACYL CARRIER (ACP) 1.  
 FT DOMAIN 2007 2074 ACYL CARRIER (ACP) 2.  
 FT DOMAIN 3045 3112 ACYL CARRIER (ACP) 3.  
 FT DOMAIN 4080 4147 ACYL CARRIER (ACP) 4.  
 FT DOMAIN 5124 5191 ACYL CARRIER (ACP) 5.  
 FT DOMAIN 6167 6234 ACYL CARRIER (ACP) 6.  
 FT BINDING 1000 1000 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 2037 2037 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 3075 3075 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 4110 4110 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 5154 5154 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 FT BINDING 6197 6197 PHOSPHOPANTHETHEINE (BY SIMILARITY).  
 SQ SEQUENCE 6486 AA; 724011 MW; 4934900AF07DF786 CRC64;

Query Match 4.1%; Score 111.5; DB 1; Length 6486;  
 Best Local Similarity 20.6%; Pred. No. 59;  
 Matches 131; Conservative 84; Mismatches 224; Indels 197; Gaps 33;

11 PSETISIMTRS---LISLCLLCSCSMPLPLEERTESR--HFTSKRPVRLDNLQIR 65  
 2586 PQPVGIMNRSEVNVGLTALIKAG-GAVVPIDPEYPEREAVVMTDCQARL--VLTK 2642  
 66 HPHHTNGLS---DIYLDNPHFAFAARAALIESAHSLSLDIYIRNDISG----- 114  
 2643 HLGAKIGSSVTAECYL--DDESNYGVHRSNIQPIITADLAIIITSGTGKPKVME 2700  
 115 ---LLENVYLAERGVRVRLLDNNTRGLDLDLALDSHPNIEVRLFPV----- 164  
 2701 HRSIYVNVLMKRAEYQMKV-----GDRSLSLSPAFDA---FVLSFPTPVLSGATVVL 2750  
 165 -----LRKWRALG---YLDPPLNRNMHNSKSTADR---ATILGGRNI----- 203  
 2751 AEDEAKDPVSLKLLIAASRCTLMTGVPSPLEALIECSPTADRIPLQVTGLGKTKAOL 2810  
 204 -----GDEYFYGEDTVPADLDI-----LAT 224  
 2811 VEKCKQLNDIYVNEGYPTSSSVATWQRLAGPDAIITIGRIANTSLYVQYHQLQP 2870  
 225 GSVMGEV---SHDFR-YW-----ASHSAHNATRIINSNGIKGL----- 260  
 2871 IGAVGEGICIGRGLRGVWNNKPALTEKEKFSHPFAAGERKYGKGLDKMLPDGTIEYIGR 2930  
 261 -----GALCYNDTS---RHALLR-----YREYVQSLY-----QKIQGRID 296  
 2931 IDEOVNVRGRRIEIGIESLLAAEKRLAIVVVYEDQJQSALAAYFTADEQLDVTKL- 2989  
 297 MOSVOTRLIS--DDPAGGLDRDRKPPAGRL--ODALKOPEKSVYLVSPVFPKSGDAL 354  
 2990 WSHLSKRLPSYMLPAHFVQDLQPLTPNGKVKDKALPKPEGKPYTERQYVAPINAVESKL 3049  
 355 AKLVQDQIDVT---VLTN-----SLQATDVAA-VHSGY-----VKYRKPLAKA--- 393

DB 3050 AEIWERVLVSGIGILDNFEQJGSHLKAMAAVAGVHREYQELPLKVLFAOPTIKALAQ 3109  
 394 -----GKILY-----ELQPNNAVPAIK-----DKGLTGSSVTSLLAKITFVDCK 432  
 DB 3110 YVATSKETEVLPLEAPALOEYYPVSSAOKRMVYLROFAPGTGVYNNPSALYIEGDL-DRK 3168  
 433 RIFIGSFNDPSPARLNTMGVATESP--KIAEOME 466  
 DB 3169 REEALHGLVERHESLRTSFHYNGEYVORVHEHVE 3204

RESULT 26  
 ID VENV VACCC STANDARD; PRT: 372 AA.  
 AC P20538; 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Major envelope protein (37 kDa protein) (p37K).  
 GN F313.  
 OS Vaccinia virus (strain Copenhagen).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;  
 OC Orthopoxvirus.  
 OX NCBI\_TaxID=10249;  
 RP [1]  
 RX MEDLINE#91021027; PubMed-2219722;  
 RA Goebel S.J., Johnson G.P., Perkins M.E., Winslow J.P.,  
 RA Paolelli E.;  
 RL "The complete DNA sequence of vaccinia virus.";  
 RL Virology 179:247-266(1990).  
 RP [2]  
 RP COMPLETE GENOME.  
 RA Goebel S.J., Johnson G.P., Perkins M.E., Davis S.W., Winslow J.P.,  
 RA Paolelli E.;  
 RL "Appendix to 'The complete DNA sequence of vaccinia virus'.";  
 RL Virology 179:517-563(1990).  
 CC -I- FUNCTION: THIS PROTEIN IS THE MAJOR ANTIGEN ON THE ENVELOPE OF  
 CC EXTRACELLULAR VACCINIA VIRUS.  
 CC -I- PTM: ACETYLATED BY PALMITIC ACID GROUP(S).  
 CC -I- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: M35027; AAA48031.1; -  
 DR PIR: I42507; MWVZCN.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 DR PROSITE: PS50035; PLD; 1.  
 KW Late protein; Antigen; Lipoprotein; Palmitate.  
 FT DOMAIN 307 334 PLD PHOSPHODIESTERASE.  
 SQ SEQUENCE 372 AA; 41823 MW; B486783DC2EFB83 CRC64;

Query Match 4.0%; Score 108.5; DB 1; Length 372;  
 Best Local Similarity 21.7%; Pred. No. 15;  
 Matches 83; Conservative 65; Mismatches 130; Indels 105; Gaps 20;

88 ARALIESAHSIDL-----QYIY-----RNDIS---GLLFLNLY 121  
 11 AKCRIVETIPENDPFRSDHLTTPCEFNELITLAKKYIYIASFCNPLSTTRGALIPDKLK 70  
 122 LAERGVAVRLLDNNTRGLDLDLALDSH--PNIEVRLFPVRLKWRALG-----YL 174  
 71 EASEGKIITVLDERKRNVLGE---LOSHCPDINFTIVN---IDKNNVGLLCCGFV 123

```

OY 175 TDPEPLNFRMKNKSTPAADNATRIILGGRNIGDEYFKVEDYFPAADLDLJGVSVGEVSHD 224
Db 124 SDDEHC--YVGNASTGSGIHTI---KTLG-----YSDPYPLAT-----DLRR 163
OY 235 FDRYWSHSAHNATRIIRSG-----NIGKGLALGYNDETSRHALLRYREVEQS 284
Db 164 FDTFKAFFNASKNSWMLNLCSAACCLPVSTAYIKRPIGVFTD--SPEHLIGSRDLDTD 221
OY 285 PLYOKIIONGRIDMOSVQORLITSDPBAKGLDERRKP--IAGRLODLAKOPEKSVUYL-- 340
Db 222 VYIDKLRBAK---ISIDIEHLAIPTTRVDGNSIYWPDIYNSITTEAIIINGVKIRLELVGN 278
OY 341 -----SPYVPPKSGTDALAKLVODGIDVYLT-----NSLQATDVAAYH----- 380
Db 279 WDKNDVYSMAATARSIDLALC--VQNDLSYKVFETIONNTKLILVDEYVHITSANEDGTHYQ 336
OY 381 -SGYVKYR--KPLLKAGIKLYE 399
Db 337 NHGVFSFNSIDKOLVSEAKKITE 359

```

ID	VENV_VACCI	STANDARD:	PRT:	372 AA.
AC	P26653:			
DT	01-AUG-1992	(Rel. 23, Created)		
DT	01-AUG-1992	(Rel. 23, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Major envelope protein (37 kDa protein) (p37K).			
GN	F11L.			
OS	Vaccinia virus (strain 1HD-J).			
OC	viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			
OX	NCBI_taxid=10251;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91251190; PubMed=2041074;			
RA	Schmutz C., Payne L.G., Gubser J., Wittek R.;			
RT	"A mutation in the gene encoding the vaccinia virus 37,000-M(r)			
RT	protein confers resistance to an inhibitor of virus envelopment and			
RT	release.";			
RL	J. Virol. 65:3435-3442(1991).			
CC	-1- FUNCTION: THIS PROTEIN IS THE MAJOR ANTIGEN ON THE ENVELOPE OF			
CC	EXTRACELLULAR VACCINIA VIRUS.			
CC	-1- PTM: ACYLATED BY PALMITIC ACID GROUP(S).			
CC	-1- MISCELLANEOUS: THE MUTATION IN POSITION 280 CONFERS RESISTANCE TO			
CC	AN INHIBITOR OF VIRUS ENVELOPMENT AND RELEASE.			
CC	-1- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL: M60412: AAA48236.1; -.			
DR	EMBL: M60413: AAA48237.1; -.			
DR	PIR: A40339; WMVZID.			
DR	InterPro: IPR001736; PLD.			
DR	Pfam: PF00614; PLDC; 2.			
DR	SMART: SM00155; PLDC; 2.			
DR	PROSITE: PS50035; PLD; 1.			
KW	late protein; Antigen; Lipoprotein; Palmitate.			
FT	DOMAIN 307 334 PLD PHOSPHODIESTERASE.			
FT	MUTAGEN 280 280 D->: INCBH RESISTANCE.			
SO	SEQUENCE 372 AA; 41805 MW; A2757EA5C60116A2 CRC64;			
Query Match	4.0%; Score 108.5; DB 1; Length 372;			
Query Local Similarity	21.7%; Pred. No. 1.5; Indels 105; Gaps 20;			
Matches	83; Conservative \ 65; Mismatches 130;			

[illegible]

RESULT 28	ENVY_VACCV	STANDARD:	PRT:	372 AA.
ID	ENVY_VACCV			
AC	P04021.			
DT	23-OCT-1986	(Rel. 02, Created)		
DT	23-OCT-1986	(Rel. 02, Last sequence update)		
DT	15-JUN-2002	(Rel. 01, Last annotation update)		
DE	Major envelope protein (37 kDa protein) (p37K).			
GN	Fl3L.			
OC	Vaccinia virus (strain WR).			
OC	Viruses: dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			
OX	NCBI_TaxID=10254;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86200431; PubMed=3701927;			
RA	Hirt P., Hiller G., Witek R.;			
RT	"localization and fine structure of a vaccinia virus gene encoding an			
RT	envelope antigen.";			
RL	J. Virol. 58:757-764(1986).			
CC	- FUNCTION: THIS PROTEIN IS THE MAJOR ANTIGEN ON THE ENVELOPE OF			
CC	EXTRACELLULAR VACCINIA VIRUS.			
CC	- PTM: ACYLATED BY PALMITIC ACID GROUP(S).			
CC	- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M12882; AAA48235.1; .			
DR	PIR; A03869; MWV237.			
DR	InterPro; IPR001736; PLD.			
DR	Pfam; PF00614; PLDC; 2.			
DR	SMART; SM00155; PLDC; 2.			
DR	PROSITE; PS50035; PLD; 1.			
FT	Late protein: Antigen; Lipoprotein; Palmitate.			
FT	DOMAIN 307 334 PLD PHOSPHODIESTERASE.			
SQ	SEQUENCE 372 AA; 41795 MW; 27AE138A382EE934 CRC64;			

Query Match 4.08; Score 108.5; DB 1; Length 372;



ID	ENTRY	STANDARD:	PRT:	372 AA.
AD	VEEV_VACCP			
AC	P29855:			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Mjor envelope protein (37 kDa protein) (p37k).			
GN	p37 OR F5.			
OS	Vaccinia virus (strain L-1VP).			
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			
OX	NCBI_TaxID=31531:			
ON	[1]			
RP	SEQUENCE FROM N.A.			
RA	Mikryukov N.N., Chizhikov V.E., Prihod'ko G.G., Urmanov I.M.,			
RA	Seplinski O.I., Blinov V.M., Nikulin A.E., Vasilenko S.K.;			
RT	"Structural-functional organization of segment of vaccinia virus			
RT	genome.";			
RL	Biotechnology 4:442-449(1988).			
CC	-1- FUNCTION: THIS PROTEIN IS THE MAJOR ANTIGEN ON THE ENVELOPE OF			
CC	EXTRACELLULAR VACCINIA VIRUS.			
CC	-1- PFM: ACTIVATED BY PALMITIC ACID GROUP(S).			
CC	-1- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION			
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -			
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS			
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY			
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL			
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <a href="http://www.isb-sib.ch/announce/isb-sib.ch">http://www.isb-sib.ch/announce/isb-sib.ch</a> ).			
CC	OR SEND AN EMAIL TO <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL, M57977: AAA48284.1: -			
DR	InterPro: IPR001736; PLD.			
DR	Pfam: PF00614; PLDC; 2.			
DR	SMART: SM00155; PLDC; 2.			
DR	PROSITE: PS50035; PLDC; 1.			
FT	Late protein; Antigen; Lipoprotein; Palmitate.			
FT	DOMAIN 307 334 PLD PHOSPHODIESTERASE.			
5Q	SEQUENCE 372 AA: 41796 MW: 2406844DPA264EF6 CCK64;			

ID	VEIN_VARY	STANDARD;	PRT;	372 AA.
AC	VEIN_VARY			
AC	P33815:			
DT	01-FEB-1994	(Rel. 28, Created)		
DT	01-FEB-1994	(Rel. 28, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Major envelope protein (37 kDa protein) (p37k).			
GN	F13L OR C17L.			
OS	Varicella virus.			
OC	Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;			
OC	Orthopoxvirus.			
OX	NCBI_TaxID=10255;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=India-1967 / Isolate Ind3;			
RX	MEDLINE=94152154; PubMed=8109158;			
RA	Shchelkunov S.N., Blinov V.M., Resenchuk S.M., Totmenin A.V.,			
RA	Sandakhchiev L.S.;			
RT	"Analysis of the nucleotide sequence of a 43 kbp segment of the			
RT	genome of varicella virus India-1967 strain.";			
RL	Virus Res. 30:239-258(1993).			
RN	[2]			
RP	COMPLETE GENOME.			
RC	STRAIN=India-1967 / Isolate Ind3;			
RA	MEDLINE=93202281; PubMed=8384129;			
RA	Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;			
RT	"Genes of varicella and vaccinia viruses necessary to overcome the host			
RT	protective mechanisms.";			
RL	FEBS Lett. 319:80-83(1993).			
CC	-I- FUNCTION: THIS PROTEIN IS THE MAJOR ANTIGEN ON THE ENVELOPE OF			
CC	EXTRACELLULAR VACCINIA VIRUS.			
CC	-I- PTM: ACYLATED BY PALMITIC ACID GROUP(S).			
CC	-I- SIMILARITY: CONTAINS 1 PLD PHOSPHODIESTERASE DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no			
CC	way			

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X69198; CAA48978.1; -  
 DR PIR: H36840; H36840. PLD.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDc: 2.  
 DR SMART: SM00155; PLDc: 2.  
 DR PROSITE: PS00035; PLD: 1.  
 DR Late Protein: Antigen: Lipoprotein: Palmitate.  
 KW DOMAIN 307 334 PLD PHOSPHODIESTERASE  
 FT SEQUENCE 372 AA: 41902 MW: C769B05D048EC944 CRC64;

Query Match 3.9%; Score 106.5; DB 1; Length 372;  
 Best Local Similarity 21.4%; Pred. No. 2.1;  
 Matches 82; Conservative 65; Mismatches 131; Indels 105; Gaps 20;

QY 88 ARAALIESHSLD-----QYIWM-----RNDIS---GRLLENLY 121  
 DB 11 ACRCLVETLPENMDRSDHLTFECFNEIITLAKKVIYASFCNPLSTRGALIFDKLK 70  
 QY 122 LAAGVRVRRLDDNNTGDLDDLALDSH-PNIEVRLNFPVLRKMRALG-----YL 174  
 DB 71 EASERGIKIYVLDGRKRNLGE-----LQSRCPDINFTVN---IDKNNVGLLGCFFV 123  
 QY 175 TDFPLNRMRHMKSFADNRATILGGRNIGDEYFKVGEDTFADLDILATGSGVEVSHD 234  
 DB 124 SDDEBC--YVGNASFTGSGIHTI---KTLG-----VYSDYPLPLAT-----DLRRR 163  
 QY 235 PDYRNASHAHNATRLIRSG-----NIGKGLQALGVNDESRHALLRYREVES 284  
 DB 164 FDTFAFNSSVKNKSWLNLSSACCLPVSTRAYHKNPGLGVFTD--SPEHLGLYSDDLTD 221  
 QY 285 PLYXRIQGRIDMOSVQTRLISDPAKGLDRDRKRP--IAGRLDALKQPEKSYLV-- 340  
 DB 222 VYIDKLRSAK--TSIDIEHLAIVPTTRVDGNSYWPDIYSIIEALIRGVKIRLVGN 278  
 QY 341 ----SPYFPTSGTDALAKLVQDIDIVYLT---NSLQATDAVAH----- 380  
 DB 279 WDKNDVYSMAATVRESLDAIC--VQNDLSVAKVFITQNNTKLLIVDEYVHTTSANFDGTHYQ 336  
 QY 381 -SGYVKYR---KPLKAGIKIYE 399  
 DB 337 NHGFVSFNSIDKQVSEAKKIFE 359

RESULT 31  
 ID NUB4\_YEAST STANDARD: PRT: 726 AA.  
 AC P52891.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Nucleoporin NUP84 (Nuclear pore protein NUP84).  
 GN NUP84 OR YD1116W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96152656; PubMed=8565072;  
 RA Sifosoglou S., Wimmer C., Rieger M., Doye V., Tekotte H.,  
 RA Weise C., Emig S., Segref A., Hurt E.C.;  
 RT "A novel complex of nucleoporins, which includes Sec13p and a Sec13p  
 RT homolog, is essential for normal nuclear pores.";  
 RL Cell 84:265-275(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX Rieger M., Mueller-Auer S., Brueckner M., Schaefer M., Wegner G.;  
 RA Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: REQUIRED FOR EFFICIENT MRNA EXPORT FROM THE NUCLEUS TO

THE CYTOPLASM AND FOR CORRECT NUCLEAR PORE BIOGENESIS.  
 -!- SUBUNIT: INTERACTS WITH NUP120, NUP85, SEC13 AND SEH1.  
 CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC -!- SIMILARITY: TO RAT NUCLEOPORIN NUP107  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X90993; CAA62479.1; -  
 DR EMBL: 274164; CAA98684.1; -  
 DR SGD: S0002274; NUP84.  
 KW Nuclear protein; Transport.  
 FT SEQUENCE 726 AA: 83635 MW: F45BAAF9976EEFE CRC64;

Query Match 3.9%; Score 104.5; DB 1; Length 726;  
 Best Local Similarity 24.0%; Pred. No. 7.7;  
 Matches 36; Conservative 27; Mismatches 54; Indels 33; Gaps 5;

QY 8 CAMPSETSIMKTRLSILCLLCCSSWLPLEERTESRHRFTSKFYRDNIITQIHT 67  
 DB 199 EAKLSDNIS-----ICMLIGIQEYLPVDTQIAEFNTQGIK----- 238  
 QY 68 PHTNGLSDIYLLN-----DPHEFAFARALIESAHSIDLOYIIRNRISGRLLFNLYL 122  
 DB 239 KHSIMRRIVYSLSQAGADPYER-AIYSYLSGAIINQEVLYQSWESDLH-----IHL 290  
 QY 123 AAEGRVVRRLDDNNTGDLDDLALDSH 152  
 DB 291 NQIQTIEINYLENNQYGTDELILPSPH 320

RESULT 32  
 ID ALA8\_ARATH STANDARD: PRT: 1189 AA.  
 AC O9LK90.  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Potential phospholipid-transporting ATPase 8 (EC 3.6.3.1)  
 DE (Aminophospholipid flippase 8).  
 GN ALA8 OR AT3G27870 OR K16N12.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Columbia;  
 RX MEDLINE=20363099; PubMed=10907853;  
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.  
 RT Sequence features of the 4,251,695 bp regions covered by 90 PL, TAC  
 RT and BAC clones.";  
 RL DNA Res. 7:217-221(2000).  
 CC -!- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY  
 CC (E1-E2 ATPASES). SUBFAMILY IV.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: AP000371; BAB02533.1; -
DR InterPro: IPR001757; ATPase_E1-E2.
DR InterPro: IPR001454; Hlgcnase/hydrolase.
DR Pfam: PF00702; Hydrolase_1.
DR PRINTS: PR00119; CATATPASE.
DR PROSITE: PS00154; ATPase_E1-E2_1.
DR HydroLase: Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
KW Magnesium; Multigene family.
FT DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 72 93 POTENTIAL.
FT TRANSMEM 94 97 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 98 120 POTENTIAL.
FT TRANSMEM 121 299 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 300 321 POTENTIAL.
FT TRANSMEM 322 358 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 359 376 POTENTIAL.
FT TRANSMEM 377 920 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 921 940 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 941 954 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 955 974 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 975 1004 POTENTIAL.
FT TRANSMEM 1005 1027 POTENTIAL.
FT TRANSMEM 1028 1040 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1041 1063 POTENTIAL.
FT TRANSMEM 1064 1069 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1070 1090 POTENTIAL.
FT TRANSMEM 1091 1107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1108 1132 POTENTIAL.
FT TRANSMEM 1133 1189 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 424 424 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 865 865 MAGNESIUM (BY SIMILARITY).
FT METAL 869 869 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1189 AA; 135309 MW; DCF951CF3FCF55E83 CRC64;

```

```

Query Match 3.9%; Score 104.5; DB 1; Length 1189;
Best Local Similarity 22.5%; Pred. No. 16;
Matches 84; Conservative 57; Mismatches 147; Indels 85; Gaps 20;

```

```

OY 151 SHPNTEVRLNFNFVLRKMRALGYL-TDPRRLNRHMKSEFA---DNRTTILGRRIGSD 206
DB 500 NOPNML-----IQKFFVLAICHTAIDVNSDGETIYEESPDEAFVLASRELGEF 553
OY 207 YKVEDIV-FADLDILATVGVGEVSHDFDRYMAHNAHTRIRSGN-----IGKGL 260
DB 554 FFSRSQTSISLIEIDHM-TGEKVDRYELHVLFESSSKRSYVRNENRLLLSKGA 612
OY 261 QALGVNDETSRLALRYRTVEOSPLYQK-IQGRIDMSVQTRISDP----- 309
DB 613 DSVMEK-RLAKHGRONERETKEIKRYAEGALRTIVTY-----REIDDEVYVWEFEFL 666
OY 310 -AKGL---DRDRKPPFIARLDALAKOPEKSYIVSPYVP---TKSGDALAKLVQDDI 362
DB 667 NAKTLVTEPRD-----ALIDAAADIKIDLLLGSTVDEKLVQGVPCIEKLSQAV 719
OY 363 DVTVLNLSQATDVAHVSGYVYKRPDLKAGIKITELQPNHAVPAKTKGILGSSVSTL 422
DB 720 KIMVLVLGDKET---AINIGYA---CSLRREGK-----QILVTLDP-----SSIEKL 761
OY 423 HAKTFLVDCGRKIFIGSFN-----LDRSRARLNTDM-GVVIIESPKIAEQNE 466
DB 762 EKQG---DKDAVAKASFOSIKQLREGMSQTAAVTNSAKENSEMGLVIDGKSLTYALD 818
OY 467 RRLADTSPEYAYR 479
DB 819 SKLEKEFLELAIR 831

```

```

RESULT 33
PDP1_ARATH STANDARD; PRT; 1096 AA.
AC 09LR25;
DT 15-JUN-2002 (Rel. 41, Created)

```

```

DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase D p1 (EC 3.1.4.4) (AtPLDp1) (phospholipase D1 PHOX and
DE PX containing domain) (phospholipase D zeta 1) (PUDZeta1).
GN PUD1 OR AT3G16785/AT3G16790 OR MGL6.24/MGL6.26/MGL6.27 OR K2019_1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Qin C., Wang X.;
RT "Arabidopsis phospholipase D zeta1."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty p1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
CC -1- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal
CC phosphodiesteric bond. Phosphatidylcholine-selective.
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatidate.
CC -1- ENZYME REGULATION: Calcium-independent and PIP2-dependent.
CC -1- SIMILARITY: Belongs to the phospholipase D family. PXPB-PLD
CC subfamily.
CC -1- SIMILARITY: CONTAINS 1 PHOX HOMOLOG (PX) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: AF411833; AAL06337.1; -
CC EMBL: AB028608; BAA95772.2; -
CC EMBL: AB022217; BAA95772.2; JOINED.
DR InterPro: IPR001849; PH.
DR InterPro: IPR001736; PLD.
DR Pfam: PF00169; PH.*.1.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00233; PH; 1.
DR SMART: SM00155; PLDC; 2.
DR SMART: SM00312; PX; 1.
DR SMART: SM00155; PX; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50035; PLD; 2.
DR PROSITE: PS50195; PX; FALSE NEG.
KW HydroLase; Lipid degradation; Repeat; Multigene family.
FT DOMAIN 50 204 PH.
FT DOMAIN 234 342 PLD.
FT DOMAIN 477 504 PH.
FT DOMAIN 892 919 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 482 482 POTENTIAL.
FT ACT_SITE 484 484 POTENTIAL.
FT ACT_SITE 489 489 POTENTIAL.
FT ACT_SITE 897 897 POTENTIAL.
FT ACT_SITE 899 899 POTENTIAL.
FT ACT_SITE 904 904 POTENTIAL.
SQ SEQUENCE 1096 AA; 124504 MW; 8B71D9DDDEA45 CRC64;

```

```

Query Match 3.8%; Score 104; DB 1; Length 1096;
Best Local Similarity 17.6%; Pred. No. 15;
Matches 107; Conservative 78; Mismatches 154; Indels 270; Gaps 25;

```

QY	81	DPEHFAARALIAASHSDL-----QYIWM--NPISGRLLEIYVLAEGVVR	131
Db	375	DGGAAFAIAIAIENAKSELFICGMYWCPRLIIRFPDPTSRDLNLENAKAGVOIT	434
QY	132	LILDDNNTRIGDLLLALDHPNIEVRLFNPFVLRKRALGLTPEPRLN-----	182
Db	435	ILI-----YKEVALALAKTINS-----YSKRRLIG-IHEVRLRPDIFFSGV	476
QY	183	---RMHKSFTADRAIILIGRNI-----GDEF-----KVG	211
Db	477	YLSHHEKLYIVDMQVCFIGGLDCGRDYTFEHKVGNDSYTWPKDKDYNPREESEPMTW	536
QY	212	EDTV-----FADLDILATGYSVYGEHDFDRYMAASHANATRI-----	250
Db	537	EDALKDELERKKPRMFMHVCALMGPPCPDVARHFVQRW-NYAKRNAPYEDSIPLLM	595
QY	251	-----INSGNIGKGLQALGYNETSRHA-----LIRREIVEQ---	283
Db	596	POHMVPIPHYMGROEESDIESKKEEDSIRGIRRDDSFSSRSSLIODIPLLPHEPYDQGS	655
QY	284	-----SPL-YOKIOTGRIDMQSVQVRLISDDPAKGL--DRDRKRPPIAGRLQ	327
Db	656	SGGKENGKTNRRNGPFSEFRASKIEPVD-----GDTPMKGFYDDRKGDLDPVAKKGS	706
QY	328	DAL-----KQPEKS-----	336
Db	707	NAIDSEMMETODHDYQVGSPEDETGQVGPRTSCRCQIIRISVQMSAGTSQVEESHSAVRS	766
QY	337	-----VILVSYFVPPTSGSD-----ALALV-----	358
Db	767	LIDKAEHFIYIENQFITSGLSGDDTVYKNRYLALRYLIRAHNEKKIFRVVVVIELLPGF	826
QY	359	QGDGID-----VYVLTASLQATVPAVAHSGVYKRPFLKAGIK	396
Db	827	QGGIDDSGAASVRAIIMWQRTIIRGHNSITLMLYTIQVKA--HXYISF-----YGLR	878
QY	397	LYELQPMHNPATKDKGLTGGSVTSLHAKTFIVDGKRIFIGSFNLDPRS--ARLTEMGV	454
Db	879	AY-----GKISEDDPVATSGQVYVHSKIIMVDDRALIGSANINDRSLGSRDSEIGV	930
QY	455	VIESPKIAE 463	
Db	931	LIEDTELVD 939	
RESULT 34			
ATL_STAU			
ID	ATL_STAU	STANDARD;	PRT: 1256 AA.
AC	P52081;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Bifunctional autolysin precursor [includes: N-acetylmuramoyl-L-alanine		
DE	amidase (EC 3.5.1.28); Mannosyl-glycoprotein endo-beta-N-		
DE	acetylglucosamidase (EC 3.2.1.96)].		
GN	ATL.		
OS	Staphylococcus aureus.		
OC	Bacteria; Firmicutes; Bacilliales; Staphylococcus.		
OX	NCBI_TaxID=1280;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 205-214 AND 776-792.		
RC	STRAIN-RNA50;		
RX	MEDLINE=95116542; Pubmed=7816834;		
RA	Oshida T., Sugai M., Komatsuzaawa H., Hong Y.-M., Sugihara H.,		
RA	Tomasa A.;		
RT	"A Staphylococcus aureus autolysin that has an N-acetyl[muramoyl]-L-		
RT	alanine amidase domain and an endo-beta-N-acetylglucosamidase		
RT	domain: cloning, sequence analysis, and characterization.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 92:285-289(1995).		
RN	[2]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-NCTC 8325-4;		

```

RA   Foster S.J.:
BL   Submitted (APR-1995) to the EMBL/Genebank/DBJ databases.
CC
CC -I- FUNCTION: ENDOHYDROLYSIS OF THE DI-N-ACETYLCHITOBIOSYL UNIT IN
      HIGH-MANNOSE GLYCOPOLYMERIDES AND GLYCOPROTEINS CONTAINING THE
      -(MANNAOSE 6)-GLUCAMIDE(2)-ASIN STRUCTURE. ONE N-ACETYLD-D-GALACTOSAMINE
      RESIDUE REMAINS ATTACHED TO THE PROTEIN; THE REST OF THE
      OLIGOSACCHARIDE IS RELEASED INTACT.
CC CATALYTIC ACTIVITY: Hydrolyzes the link between N-acetylmuramoyl
      residues and L-amino acid residues in certain bacterial cell-wall
      glycopeptides.
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of the di-n-acetylchitobiosyl
      unit in high-mannose glycoproteins and glycoproteins containing
      the -(Man(GlcNAc)(2)Asn)-structure. One N-acetyl-D-glucosamine
      residue remains attached to the protein; the rest of the
      oligosaccharide is released intact.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- PM: UNDERGOES PROPEPTOLYTIC PROCESSING TO GENERATE THE TWO
      EXTRACELLULAR LYTC ENZYMES.
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE N-
      ACETYLMURAMOYL-L-ALANINE AMIDASE FAMILY 2.
CC -I- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO FAMILY 73 OF
      GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D17366; BA004185.1; -
DR EMBL; L41499; AA89982.1; -
DR InterPro; IPR002502; Amidase_2.
DR InterPro; IPR002901; Amidase_4.
DR Pfam; PF01510; Amidase_2; 1.
DR Pfam; PF01832; Amidase_4; 1.
DR SMART; SMO0047; Lyz2; 1.
KW Cell wall, Hydrolase, Signal, Multifunctional enzyme, Repeat.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 1256 BIFUNCTIONAL AUTOLYSIN.
FT DOMAIN 199 775 N-ACETYLMURAMOYL-L-ALANINE AMIDASE.
FT DOMAIN 776 1256 ENDO-BETA-N-ACETYLGALACTOSAMIDASE.
FT REPEAT 425 589 1.
FT REPEAT 596 758 2.
FT REPEAT 770 932 3.
SQ SEQUENCE 1256 AA; 137384 MW; 2BB76CAA92FDD20 CRC64;

Query Match          3.8%; Score 103.5; DB 1; Length 1256;
Best Local Similarity 19.5%; Pred. No. 20;
Matches 101; Conservative 61; Mismatches 140; Indels 215; Gaps 26.

OY    4 2 EERTSRHFNSTKPYRLNIIQIRHT-PHTNGLSDIYLLNDPHEAFAPAALIESAEHS- 99
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     769 QPKTVAKVAYTKPPQTQTSKRIAOVKNMNGCI-----RASVEKTAKNG 813
               : : : : : : : : : : : : : : : : : : : : : : : :
OY    100 ---IDLQYYIMRNDSGRLEFLNYLAERGVARYALLDDNNTRQLDLLDLASHPNIE 156
              : : : : : : : : : : : : : : : : : : : : : : : :
Db     814 AKYAARTRYTYVKERAAGNETYVL-----NNT-----SH-NIP 845
              : : : : : : : : : : : : : : : : : : : : : : : :
OY    157 VLFNPFWLRKKRALG---YLTDFPLNRMRMNKSFT--ADRRATILGGRGINDGEYF-- 208
              : : : : : : : : : : : : : : : : : : : : : : : :
Db     846 LGWFFWKDLNV-QNLGKEKTKTKTKVAKSNNGSLMVPMWGTFNQVIILGNNTAOGTFMAT 904
              : : : : : : : : : : : : : : : : : : : : : : : :
OY    209 ---KYGEDT-VFADL-----DILTATSGVGEEVSHDPRIYASHAHNATRPIISG 254
              : : : : : : : : : : : : : : : : : : : : : : : :
Db     905 KOVSXGKVUYLVGTINNRTGWVANAKDLTAPlAVKFTTSAKD-----YVTYIVIKNG 956
              : : : : : : : : : : : : : : : : : : : : : : : :
OY    255 NIGKVLQALGYDETSSHALLYRE----TYQSPF-----YKIQTGRIQDMQSVQRU 304
              : : : : : : : : : : : : : : : : : : : : : : : :
Db     957 N---GIYVVTPNSDTAKTSLKAFAEQPAVYVKEQVINGOTWTYGALSNGKLAW-----I 1007
              : : : : : : : : : : : : : : : : : : : : : : : :
OY    305 ISDDPAK-----GLDDRORRPPIAGRLODA----- 329
```

```

Db 1008 KSTDLAKELIKYNGTGMTLNOVAQIOAGLOPKYQVORVPGKMTAKFENDVKHAMDTKRLA 1067
Oy 330 -----LKQPE-----KSYLYSPRYV 345
Db 1068 ODPALKYQFLRDOPONISIDKINQFLKGVLENQGAENKAQMYGINEYLLISHALL 1127
Oy 346 PTKSGTDLAKLAVQDIDVT---VLJNS-----LQATDVAAVHSGYVYKXK----- 388
Db 1128 EYNGNTSQLAK-----GADVNNKVVYNTNSTKYNHVGIAIYNDPLREG-IXYAKQAGMD 1182
Oy 389 -----PLKAGIK-LYELOPNHVPAT 409
Db 1183 TVSKAIVGAKFIGNSYVAKGNTLYKKMRMNAHPGT 1219

RESULT 35
PLD1_MOUSE STANDARD: PRT: 1074 AA.
ID PLD1_MOUSE
AC 092280: 035911;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase D1 (EC 3.1.4.4) (PLD1) (Choline phosphatase 1)
DE (Phosphatidylcholine-hydrolyzing phospholipase D1) (mPLD1).
GN PLD1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS PLD1A AND PLD1B).
RC STRAIN=129; TISSUE=Embryonic brain, and Neonatal brain;
RX MEDLINE=97439716; PubMed=9307024;
RA Colley W.C., Altschuller Y.M., Sue-Ling C.K., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Branch K.D., Tsirka S.E., Bollag R.J.,
RA Bollag W.B., Frohman M.A.;
RT "Cloning and expression analysis of murine phospholipase D1.";
RL Biochem. J. 326:745-753(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM PLD1A).
RC STRAIN=129;
RX MEDLINE=99033000; PubMed=9813240;
RA Redina O.E., Frohman M.A.;
RT "Genomic analysis of murine phospholipase D1 and comparison to
RT phospholipase D2 reveals an unusual difference in gene size.";
RL Gene 222:53-60(1998).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=97199446; PubMed=9395408;
RA Colley W.C., Sung T.-C., Roll R., Jenco J.M., Hammond S.M.,
RA Altschuller Y.M., Bar-Sagi D., Morris A.J., Frohman M.A.;
RT "Phospholipase D2, a distinct phospholipase D isoform with novel
RT regulatory properties that provokes cytoskeletal reorganization.";
RL Curr. Biol. 7:191-201(1997).
RN [4]
RP FUNCTION: IMPLICATED AS A CRITICAL STEP IN NUMEROUS CELLULAR
RN PATHWAYS, INCLUDING SIGNAL TRANSDUCTION, MEMBRANE TRAFFICKING, AND
RN THE REGULATION OF MITOSIS. MAY BE INVOLVED IN THE REGULATION OF
RN PERINUCLEAR INTRAVESICULAR MEMBRANE TRAFFIC.
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatidate.
CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-
CC BISPHOSPHATE AND PHOSPHATIDYLINOSITOL 3,4,5-TRISPHOSPHATE,
CC ACTIVATED BY THE PHOSPHOKINASE C-ALPHA, BY THE ADP-RIBOSYLATION
CC FACTOR-1 (ARF-1), AND IN A LESSER EXTENT BY GTP-BINDING PROTEINS:
CC RHO A, RAC-1 AND CDC42.
CC -1- SUBCELLULAR LOCATION: PERINUCLEAR REGIONS: ENDOPLASMIC RETICULUM,
CC GOLGI APPARATUS AND LATE ENDOSOMES.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: PLD1A (SHOWN HERE) AND PLD1B;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY, LUNG, AND AT A MUCH LOWER
CC LEVELS, IN BRAIN, LIVER, HEART, TESTIS, AND SPLEEN.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED MOST STRIKINGLY IN SELECTED

```

```

CC CC VENTRICULAR CELLS LINING THE SPINAL CORD AND BRAIN. THE LEVEL OF
CC CC EXPRESSION DECREASES DRAMATICALLY AS THE CELLS DIFFERENTIATE INTO
CC CC NEURONS AND MIGRATE TO THE OUTER LAYER OF THE SPINAL CORD AND
CC CC BRAIN. EXPRESSION IS OBSERVED DURING DEVELOPMENT IN A RESTRICTED
CC CC REGION OF THE NASAL NEUROEPITHELIUM.
CC CC -1- DISEASE: Defects in PLD1 may result in coa which is associated
CC CC with coat color dilution and white spotting. It is also
CC CC associated with platelet-storage pool deficiency characterized by
CC CC decreased levels in serotonin and dense granules.
CC CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
CC CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@sib-sib.ch).
CC CC -----
DR DR EMBL: U87868; AAB81245.1; -
DR DR EMBL: AF083497; AAC84041.1; JOINED.
DR DR EMBL: AF083475; AAC84041.1; JOINED.
DR DR EMBL: AF083476; AAC84041.1; JOINED.
DR DR EMBL: AF083478; AAC84041.1; JOINED.
DR DR EMBL: AF083479; AAC84041.1; JOINED.
DR DR EMBL: AF083480; AAC84041.1; JOINED.
DR DR EMBL: AF083481; AAC84041.1; JOINED.
DR DR EMBL: AF083483; AAC84041.1; JOINED.
DR DR EMBL: AF083484; AAC84041.1; JOINED.
DR DR EMBL: AF083485; AAC84041.1; JOINED.
DR DR EMBL: AF083486; AAC84041.1; JOINED.
DR DR EMBL: AF083488; AAC84041.1; JOINED.
DR DR EMBL: AF083489; AAC84041.1; JOINED.
DR DR EMBL: AF083490; AAC84041.1; JOINED.
DR DR EMBL: AF083492; AAC84041.1; JOINED.
DR DR EMBL: AF083494; AAC84041.1; JOINED.
DR DR EMBL: AF083495; AAC84041.1; JOINED.
DR DR EMBL: AF083496; AAC84041.1; JOINED.
DR DR MGD: MGI:109585; PLD1.
DR DR InterPro: IPR001849; PH.
DR DR InterPro: IPR001736; PLD.
DR DR InterPro: IPR001683; PX.
DR DR Pfam: PF00169; PH: 1.
DR DR Pfam: PF00614; PLDC: 2.
DR DR Pfam: PF00787; PX: 1.
DR DR SMART: SM00233; PH: 1.
DR DR SMART: SM00155; PLDC: 2.
DR DR SMART: SM00312; PX: 1.
DR DR PROSITE: PS50003; PH.DOMAIN: 1.
DR DR PROSITE: PS50035; PLD: 2.
DR DR PROSITE: PS50195; PX: 1.
KW Hydrolase; Lipid degradation; Membrane; Alternative splicing;
KW Endoplasmic reticulum; Golgi stack; Repeat.
FT FT DOMAIN 219 328 PH.
FT FT DOMAIN 81 212 PH.
FT FT DOMAIN 459 486 PLD PHOSPHODIESTERASE 1.
FT FT DOMAIN 891 918 PLD PHOSPHODIESTERASE 2.
FT FT DOMAIN 463 928 CATALYTIC.
FT FT VARSPLC 585 623 CATALYTIC.
FT FT CONFLICT 71 74 -> N (IN ISOFORM PLD1B).
FT FT CONFLICT 574 574 NIOT -> OHP (IN REF. 1).
FT FT CONFLICT 675 675 A -> G (IN REF. 1).
FT FT CONFLICT 781 781 R -> I (IN REF. 1).
FT FT CONFLICT 876 876 C -> V (IN REF. 1).
FT FT CONFLICT 876 876 N -> V (IN REF. 1).
FT FT CONFLICT 990 990 T -> A (IN REF. 1).
FT FT CONFLICT 1071 1073 EWM -> SLT (IN REF. 1).
SQ SEQUENCE 1074 AA; 123966 MW; E11260982A217280 CRC64;

```

Query Match

3.8%;

Score 103; DB 1; Length 1074;

Best Local Similarity 20.58; Pred. No. 17;  
Matches 131; Conservative 79; Mismatches 174; Indels 256; Gaps 39;

```

QY 60 NILQIRHTPHNGSLDYL-----LNDPEFAFAARALIESAHSGLDQYIWRND- 110
D 444 NIKWRHHDHNS--SSVLMHAHEKLYIIDOSVAVG-----GIDLAYGWDDNE 491
QY 111 -----ISGRLLFNLYLAERGVRRL-----LDD--DNNRGL 142
D 492 HRLDVGSVKRTSGLSIGSLTAAVESMESLTKDKHEFHKKKEPISKIVDETDKLGKI 551
QY 143 D-----DILLALD-----SHIPNL-----EVLNFNFPV 164
D 552 GKSRKESKFSLYROLHRLHHLNADSISSIDTSYFSCRSQNLHLGLKPLKLFHPSS 611
QY 165 LKRWALGLYLDLFPRLNRHMKSFADNRATILG-----GRNIDEXFK--V 210
D 612 ESEGLTGHNSD-----TGSIRSVQVGVGLHGETRPMWCKDCNPFKRW 658
QY 211 GEDVVFAD-LDILAT-----GSV--GEVSHDFDRYASHAHNAATRIISGNIGKL 260
D 659 QLDPEFADIFIRYSPRPMWHDIGSVHGAARVARRHFIOR--WNFTKIMK-----PKY 711
QY 261 QALGY-----NDETSRHALLRYREVEESPLXOKTQGRIDM-----QSVOTRL-- 305
D 712 KSLSTPFLPKRSQATAHE-LRT-QVPGAVPAKVOLLRSADMSAGIKHHESTHAAHYTHV 769
QY 306 -----SDDPA--KGLDRRKRPPIAGLQDALOKPEK-SVYVSPY 343
D 770 IENSKHYIYIENOFISACDDKVFYFNKYSR-----IAQNLKMRHRCQRYRYIVP- 822
QY 344 FVPTKSGTDALAKLVQSDIVYVLT--NSIQADIDVAHSGYKRPPLKAGIKLYEL 400
D 823 -----LLPFGEDISITGGGNALQAI-----MHFNRYFMCGESSILPL 861
QY 401 QP-----NHAVPATKDKGLGSSVTSL--HAKFTYDGRIRFIQSNL 441
D 862 KPELGKKNINISFCGLRTHA-----ELGONLVTEIYHSHKLLADNDTVIIGSNTI 914
QY 442 DPRS--ARLNTMGVYIES--PKIAB--QMERLTADTSPEYAVVTL--DRNR 487
D 915 NDRSKLGRDSEMAVIYQDTEVPVMDGKEVQAGRFAARDLRL-CPFLVGYLSDPED 973
QY 488 LQMHDPARTKTYPNPEPKLKRILAA-----KILSLIP 520
D 974 LQ--DPVSDKEF-----KEIYVSTAAARNATYDKVFRCLP 1006

RESULT 36
PDZL_ARATH STANDARD: PRT: 820 AA.
AC PS8766;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phospholipase D zeta (EC 3.1.4.4) (AtPLDzeta) (PLD zeta).
GN PLDZETA OR AT5G25370 OR F18G18.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Columbia;
RX MEDLINE=21016721; PubMed=11130714;
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asanizu E.,
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
RA Nakazaki N., Natsu K., Okumura S., Shino S., Takeuchi C., Wada T.,
RA Watanabe E., Yamada M., Yasuda M., Sato S., de la Bastide M.,
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
RA Habermann K., Murray J., Johnson D., Kohling T., Neilson J.,

```

```

RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
RA Butler E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,
RA Wagner-Meherson C., Woliam A., Yokum M., Bell M., Dedhia N.,
RA Parnell L., Shan R., Rodriguez M., Hoon See L., Vil D., Baker J.,
RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Maria M.,
RA Mattiensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
RA Volckaert G., Wambutt R., Duesterhoeft A., Stekema W., Pohl T.,
RA Entian K.-D., Terryn N., Harley N., Bent E., Johnson S.,
RA Langham S.-A., McCullagh B., Robben J., Grymoprez B., Zimmermann W.,
RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
RA van Staveren M., Dirkse W., Moollman P., Klein Lankhorst R.,
RA Weitzensager T., Bothe G., Rose M., Haut J., Bernieris S., Hempel S.,
RA Feldpausch M., Lamberth S., Villarroel R., Glalen J., Ardiles W.,
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
RA Schueller C., Zaccaria P., Mewes H.-W., Beyer M., Franz P.,
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
RT thaliana."
RL Nature 408:823-826(2000).
CC -1- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal
CC phosphodiesteric bond.
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O -> choline + a
CC phosphate.
CC -1- COFACTOR: Calcium (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (By
CC similarity).
CC -1- DOMAIN: C2 domain is a calcium-binding fold, and the binding
CC toward calcium can be anticipated for PLD alpha due to the absence
CC of two potential calcium ligands.
CC -1- SIMILARITY: Belongs to the phospholipase D family. C2-PLD
CC subfamily.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL; AC006258; -; NOT_ANNOTATED_CDS.
DR PROSITE; PS50004; C2_DOMAIN_2; FALSE_NGCG.
DR PROSITE; PS50035; PLD; 2.
KW Hydrolyase; Lipid degradation; Calcium; Repeat; Multigene family.
FT DOMAIN 1 117 C2 DOMAIN
FT FT 334 371 PLD PHOSPHODIESTERASE 1.
FT FT 662 689 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 339 339 POTENTIAL.
FT ACT_SITE 341 341 POTENTIAL.
FT ACT_SITE 346 346 POTENTIAL.
FT ACT_SITE 657 657 POTENTIAL.
FT ACT_SITE 669 669 POTENTIAL.
FT ACT_SITE 674 674 POTENTIAL.
SQ SEQUENCE 820 AA; 93362 MW; FFB9247276BCB7B CRC64;

Query Match 3.88; Score 102; DB 1; Length 820;
Best Local Similarity 18.48; Pred. No. 14;
Matches 105; Conservative 69; Mismatches 182; Indels 216; Gaps 25;

```

Db 320 PUNGSIVOGFEVATMFTH-----HOKTIYVSEVDGSLTKRRIVSFLGIDLCDR 371  
 QY 208 FKVGEDTVEA-----DIDLATGVSVEGSHDPR 237  
 Db 372 YDTEHPLEGLTNSVANDHOPNEDGASIKKGPREPHWDHCKIDGPAAMDVLNFDQ 431  
 QY 238 YWASHAHNATRIIRSGNIGKGLQALGYNETSRHALLRYE---TVESQSPLYOKIQTG 293  
 Db 432 RMMKQ-----GSG-----RRYLISMAQLAELTVPPLPIVQ--PDN 464  
 QY 294 RIDWQSVQTRLISDDPAKLDPRKPP---IAGR-----LQD---ALKQPEKSYV 338  
 Db 465 EBGWTVQVFRSIDGAVEGFPEDPREASIGLSKDNVIERSIDQAYNAIRAKNFYI 524  
 QY 339 LVSPEF-----VPTKSGTDLAKKL-----357  
 Db 525 IENQIFLGSSFGWNSHDINLAINMALQILPKELSLKIVSKIEAGERFSYIYIPLMECK 584  
 QY 358 -----VODGIDVTYLTNSLQATDVAAVHSGYVYKRPKLKAGIKLY-----398  
 Db 585 PGSASVQALIDMQRRIMEMMYTDI-----IALRKKGIDANRDYDLTFECLGNREKGV 638  
 QY 399 -ELQPNHNAVPAKDKGLTSS---VTSIAKTFIYDGKRIFTGSENLDPRS--ARLNT- 451  
 Db 639 GYLPEPEKPEANSDYARAQESRRFMIVHSHKMIYDDEYIIIGSANINQSMGGRDTI 698  
 QY 452 -MGVIESPKIAEMERTLADTSPETAYRVTL 482  
 Db 699 AMGATQPSHLSTNNMRPYQI---FSFRISL 727

## RESULT 37

FLID\_TREMA STANDARD: PRT: 673 AA.  
 AC Q9KMW7:

Db 16-OCT-2001 (Rel. 40, Created)  
 Dt 16-OCT-2001 (Rel. 40, Last sequence update)  
 Dr 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Flagellar hook-associated protein 2 (HAP2) (Flagament cap protein)  
 GN (Flagellar cap protein).

OS Treponema maltophilum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 OX NCBI\_Taxid=511160;

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 51939 / BR.

RX MEDLINE=20170682; PubMed=10708388;

RA Heuner K., Grosse K., Schade R., Goebel U.B.;

RT "A flagellar gene cluster from the oral spirochaete Treponema

malophilum.";

RL Microbiology 146:497-507(2000).

- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF  
 THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE  
 FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPING  
 STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH  
 THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT  
 POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).  
 CC - SUBUNIT: HOMOPENTAMER (BY SIMILARITY).  
 CC - SUBCELLULAR LOCATION: Periplasmic flagellum.  
 CC - SIMILARITY: BELONGS TO THE FLID FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: Y18889; CAB67252.1; -  
 DR InterPro: IPR003481; FLID.  
 DR Pfam: PF02465; FLID; 1.

KW Flagella; Periplasmic; Coiled coil.  
 FT DOMAIN 611 664 COILED COIL (POTENTIAL).  
 SQ SEQUENCE 673 AA; 73807 MW; CAD4017940B1340B CRC64;

Query Match 3.7%; Score 101.5; DB 1; Length 673;  
 Best Local Similarity 20.8%; Pred. No. 11;  
 Matches 82; Conservative 56; Mismatches 137; Indels 119; Gaps 17;

QY 154 NIEVRLFNPEVYLRKKNALGLTDF-PRLNRRMKNKSTADNRATILIGRIGDEYKVG 212  
 Db 125 NIEYAVNDKKVSNRWKG-GKLSDFVALNRRSANTL---KASITG-----ISK 168  
 QY 213 DTYEADDLIATGS---VGEVSHDFDRYWASHAHNATRIIRSGNIGKGLQ-----LG 264  
 Db 169 DKALLIESLITGSENRILTFSGAARDF-----ALQTMIGKAPAAADSF 214  
 QY 265 YNDETSRHALLRYETVE-----OSPLYOKIQGRIDWQSVQTRLISDDPA 310  
 Db 215 IKRDTIRNADNLNLSKTVFEFSQDALVPPRGFESPIQKIAEKTKTIRFNVAI-DIPA 273  
 QY 311 KGLDRRRKRPINGRL-----QDALQPEKSYVYSPYVPTKSGT 351  
 Db 274 EEVLPRKPELPSAGTVSEFKDVTLSNEEPDTKLPLSETEKVPAPVEDYAAYVKAADGT 333  
 QY 352 D-----ALAKLYODGIDVTYLTNS-----LQATDVAAVHSGYVY 386  
 Db 334 EIRPLGLKRGENTSYSTIASADYPGMSALYIKNNHTHKIKVMTKRPETIDMAA-SSGIVPL 392  
 QY 387 RKPLKAGIKLYELQPNHNAVPAKDKGLTSSVTSIAKTF-----FYDKR----- 433  
 Db 393 NPAETADAKIKKYQGITITRPSNITIDVVPVNLINQAKTEKPATISIEPDKAKAREALI 452  
 QY 434 IFIGSFNLDPRSARLTEMGVIES-PKIAEQE 466  
 Db 453 TVYKTN-----RLMAELNLTQTQPEITTELE 480

## RESULT 38

PDAL\_ORYSA STANDARD: PRT: 812 AA.  
 ID PDAL\_ORYSA

AC Q43007;  
 Dt 01-NOV-1997 (Rel. 35, Created)  
 Dt 01-NOV-1997 (Rel. 35, Last sequence update)  
 Dr 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phospholipase D alpha 1 precursor (EC 3.1.4.4) (PLD alpha 1) (Choline  
 phosphatase 1) (Phosphatidylcholine-hydrolyzing phospholipase D 1).  
 GN PLD1.

OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 OX NCBI\_Taxid=4530;

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RP STRAIN=cv. Koshihikari;

RX MEDLINE=96012933; PubMed=7551587;

RA Ueki J., Morioka S., Komari T., Kunushiro T.;

RT "Purification and characterization of phospholipase D (PLD) from rice

(Oryza sativa L.) and cloning of cDNA for PLD from rice and maize

(Zea mays L.).";

RT Plant Cell Physiol. 36:903-914(1995).

RL Plant Cell Physiol. 36:903-914(1995).

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Koshihikari; TISSUE=leaf;

RA Morioka S., Ueki J., Komari T.;

RT "Characterization of two distinctive genomic clones for phospholipase

D from rice.";

RL (in) Plant Gene Register PCR97-076.

- FUNCTION: Hydrolyzes glycerol-phospholipids at the terminal  
 phosphodiesteric bond. Plays an important role in various cellular  
 processes. Thermostable between 4 and 37 degrees Celsius, but  
 activity is reduced to one third of the original level after  
 incubation at 50 degrees Celsius for 30 minutes. Optimal pH is

```

CC 6.0.
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O = choline + a
CC phosphatide.
CC -1- COFACTOR: Calcium.
CC -1- SUBUNIT: Monomer.
CC -1- TISSUE SPECIFICITY: Expressed in leaves, roots, developing seeds
CC and cultured cells.
CC -1- DEVELOPMENTAL STAGE: The transcript begins to emerged in seeds as
CC early as the second day after imbibition and increased after
CC radicle emergence on the third day. Strongly expressed in the
CC leaves, roots and residual grain of seedling eight days after
CC imbibition. Undetectable in mature dry seeds.
CC -1- DOMAIN: C2 domain is a calcium-binding fold, and the binding
CC promotes the protein association with membranes. A lower affinity
CC toward calcium can be anticipated for PLD alpha due to the absence
CC of two potential calcium ligands.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. C2-PLD
CC SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D73411: BAA1136.1;
DR EMBL: AB001920: BAA19467.1;
DR InterPro: IPR000008: C2.
DR InterPro: IPR001736: PLD.
DR Pfam: PF00614: PLDC; 2.
DR SMART: SM00239; C2; 1.
DR SMART: SM00155: PLDC; 2.
DR PROSITE: PS00004: C2_DOMAIN_2; 1.
DR PROSITE: PS50035: PLD; 2.
KW Hydrolyase; Lipid degradation; Calcium; Repeat.
FT PROPEP 1 46
FT CHAIN 47 812 PHOSPHOLIPASE D ALPHA 1.
FT DOMAIN 1 114 C2 DOMAIN.
FT DOMAIN 330 368 PLD PHOSPHODIESTERASE 1.
FT DOMAIN 658 685 PLD PHOSPHODIESTERASE 2.
FT ACT_SITE 335 335 POTENTIAL.
FT ACT_SITE 337 337 POTENTIAL.
FT ACT_SITE 342 342 POTENTIAL.
FT ACT_SITE 663 663 POTENTIAL.
FT ACT_SITE 665 665 POTENTIAL.
FT ACT_SITE 670 670 POTENTIAL.
SQ SEQUENCE 812 AA: 92209 MW: 635F94E56CDEDF8F CRC64:
Query Match 3.7%; Score 101; DB 1; Length 812;
Best Local Similarity 20.4%; Pred. NO. 16;
Matches 108; Conservative 65; Mismatches 182; Indels 174; Gaps 26;

```

```

DB 403 PREPMDHDSRLGPIAMDVLYNFEQRMKGGCKDLLQLRDLSTIIPSPV--MEPED 460
QY 294 RIDMOSVOTRLLSDPAKALDIDRRKPPAG-----RLD-----ALKQEPKSY 338
DB 461 RETMNVQLEFRSIDGAAFGPDPTEPAKAGLVSKQDIDRSDAYIHAIIRAKNFY 520
QY 339 LVSPYF-----VPTKSGDALAKL----- 357
DB 521 IENQVFLSSYAMKPEGIKPEDIGALHLIRKLAKVSKIIAGERFYVYVPMWPECV 580
QY 358 -----VQSGID-----VTVLNLSQATDVAAVHSGYKY----RKPLKAGIKIYE 399
DB 581 PEGSVOALIDMQRRTMEMMYTDITFALQAKGIEANPKDYLFPCLGNEVQAG-----E 636
QY 400 LQPNHNAVPTKDKGLTGS-----VTSLHAKTFYDCKRIFGSPNIDPRS 445
DB 637 YQPEQPEADTDYSRAQEARREMYVHTKMTIVDDYIIIGSANINORS 685

RESULT 39
YQFR_BACSU
ID YQFR_BACSU STANDARD: PRT: 438 AA.
AC P34475.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable RNA helicase yqfr.
GN YQFR.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
(1)
SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes."
RL Microbiology 142:3103-3111(1996).
[2]
SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchi S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itoya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaubert-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mamel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Nodack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
RA Parson E., Pohl T.M., Portetelle D., Porvolik S., Prescott A.M.,
RA Preece N., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroter R., Scofield F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Totsuchi M., Tamakoshi A., Tanaka T., Terpsita P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretoli A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,

```



RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT subtilis".  
 RL Nature 390:249-256(1997).  
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D84432; BAA12495.1; -  
 DR EMBL: 299116; CAB14444.1; -  
 DR HSP: 058063; 1H8.  
 DR Subtilisin; Bg11664; ygfR.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000629; DEAD\_box.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00270; DEAD.1  
 DR Pfam: PF00271; helicase\_C.1.  
 DR SMART; SM00487; DEXDC.1.  
 DR SMART; SM00490; HELIC; 1.  
 DR PROSITE; PS00039; DEAD.ATP\_HELICASE; FALSE\_NEG.  
 DR Hypothetical protein; Helicase; ATP-binding; RNA-binding;  
 KM Complete proteome.  
 FT NP\_BIND 48 55 ATP (POTENTIAL).  
 FT SITE 156 159 DEAD BOX.  
 SQ SEQUENCE 438 AA; 50027 MW; B6D425EF6D433789 CRC64;  
 Query Match 3.7%; Score 100.5; DB 1; Length 438;  
 Best Local Similarity 22.8%; Pred. 7.2;  
 Matches 65; Conservative 32; Mismatches 83; Indels 105; Gaps 13;  
 QY 101 DLQYIWRNDISGR-----LENDVYLAEE---RGVAVRLLLDDNNRGLDILLAL 149  
 DB 180 DLQMLVFSAITPEKTLPLKKYENPKYAVHEPKQYTAKEIHLIPSKHRDKLLFDI 239  
 QY 150 DSNPNEVRLNPFVLRKRRLALCYLTFPLRNRMHNSKSTADNRATILGGRITGEYER 209  
 DB 240 MSHLN-----PYLGIVFANTKMTADHIAQYLGKGM-----K 271  
 QY 210 VG-----EDTVFADLDILATGSGVGEVSH-----DEPRYMA 240  
 DB 272 IGLHGGTLPRERKKYKWKQINDLEFYIATDLAAGIDIKGVSHVINTPELDLPFY-- 329  
 QY 241 SHSANAATRIIRSGNIGKGLQALGVNDETSRHALLRYRET-VBOSPLYOKIQTRIDWOS 299  
 DB 330 ---VHVVGRTARAGSSG---QAMTIYELDDEDLVLEKKGIE---FEYLELEKGEWK- 378  
 QY 300 VQTRLISDPAKGLDRRK-----PTIAGRLQDALDKPEK 335  
 DB 379 -----KGDROQRKKRKKRKTPEADEIAHRL---VKRPEK 409  
 RESULT 40  
 C6UL\_DROME  
 ID C6UL\_DROME STANDARD; PRT; 485 AA.  
 AC Q9V979;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 16-OCT-2001 (Rel. 40; Last annotation update)  
 DE Probable cytochrome P450 6ul (EC 1.14.-.-) (CYPI1U1).  
 GN CYP6U1 OR CG3567.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandal D., Bolshakov S.,  
 RA Borika D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,  
 RA Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos R., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacieb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP CONCEPTUAL TRANSLATION.  
 RA Nelson B.;  
 RL Unpublished observations (SEP-2000).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE METABOLISM OF INSECT HORMONES AND  
 CC IN THE BREAKDOWN OF SYNTHETIC INSECTICIDES (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 CC oxidized flavoprotein + H(2)O.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum  
 CC (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE003790; AAF57417.1; ALT\_SEQ.  
 DR FlyBase: FBgn0033121; Cyp6u1.  
 DR InterPro: IPR001128; Cytochrome\_P450.  
 DR Pfam: PF00067; P450; 1.  
 DR PRINTS; PR00385; P450.  
 DR PROSITE; PS00086; CYTOCHROME\_P450; 1.  
 DR Oxidoreductase; Monooxygenase; Membrane; Heme; Microsome;  
 KM Endoplasmic reticulum; Hypothetical protein.  
 FT BINDING 427 427 HEME (BY SIMILARITY).  
 SQ SEQUENCE 485 AA; 55206 MW; 3EC4D83BE3676AB3 CRC64;

Query Match 3.7%; Score 100.5; DB 1; Length 485;  
Best Local Similarity 20.0%; Pred. No. 8.3; Indels 181; Gaps 24;  
Matches 96; Conservative 59; Mismatches 145; Indels 181; Gaps 24;

```
OY 65 RHTPTNGLSDIYLLNDPHEAFARAALIESAHSIDLOYYIWRNDISGRLEFNLYLAA 124
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 51 RHA--QDALQDIYTAVKGRAPFGVACLKPFILADL-----KLVHQIIFTD 97
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 125 ---ERGVAVRLLDNNTRGLDLDLALDSHPNIEVRLNPFVLRKRALGYLTDPR 180
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 98 GFTSRG-----LYSNPGEPLSHNLQLDGH-----KWRSL----- 129
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 181 NRMHNS---FTADNRATILG-----GRNIGD---EYFKGEDVFPADLDILAG 225
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 ---HAKSAEVTTPANMOKLVLRLSQISSRIQIDGKSLQITINISELVGAVTDMASM 185
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 226 S--VVEVSHDFDR---YMAASHAHNATRIIRSGNIGKQALGYNDETSRHALLRYE 279
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 186 AFGLVGQDNVEFAKWTBNYWADEFBWMQAYLALFPLIAR-----LLQYKS 230
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 280 TVEQSPLY-QKIQTGRI-----DMQSVOTRLISDDPAKGLDRRKP---PIAGR--- 325
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 YAEPAIAYFOKVALSOLQHRRRDRQPIQTFL-----QLYSNAEKPLTDIEIAGQAFG 283
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 326 -----LQDALKQPEKSYLVSPYFVPTKSGTDALAKLVODGIDVTVLN 369
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 FYVLAGLPLNATLACLYELARQPE-----VQD-----RT 313
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 370 SLQATDVAAVHSGYV-----KYRKPLKAGIKLYELQPNHNAVATKDKGLTGSSVTS 421
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 RLEINKALEEHGGOVTPCELRRLRYTKOVLNETHLPHPLLRATKEFEVPGSVFY- 372
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 422 LHAKEFIYDGRIFT--GSFNLDPRSARLNTKMGVIESPK--IAEQMERTLADTSPEYA 477
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 ---IAKGNVLIPTAIIHMDPG-----IYENPQRFYPERFEQARRSRPAAA 416
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
OY 478 Y 478
   || : ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 F 417
```

Search completed: May 12, 2003, 09:58:00  
Job time : 41 secs

GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:56:11 ; Search time 30 Seconds  
(without alignments)  
514,901 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 2713  
Sequence: 1 MRANKTOAMPSETISLMKT.....KLMKRIAKIISLPRESIL 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/pdata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/pdata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/pdata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/pdata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/pdata/1/1aa/CTUS.COMB.pep:\*  
6: /cgn2\_6/pdata/1/1aa/Dackfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	295	10.9	495	US-09-134-001C-4273	Sequence 4273, Ap
2	264.5	9.7	494	US-09-134-001C-4176	Sequence 4176, Ap
3	145.5	5.4	933	US-09-141-206-2	Sequence 2, Appl
4	136	5.0	922	US-09-141-206-6	Sequence 6, Appl
5	133.5	4.9	933	US-09-107-149-17	Sequence 17, Appl
6	130.5	4.8	932	US-08-968-752B-6	Sequence 6, Appl
7	130.5	4.8	932	US-09-536-224-6	Sequence 6, Appl
8	121.5	4.5	933	US-09-107-149-2	Sequence 2, Appl
9	111	4.1	1036	US-08-968-752B-4	Sequence 4, Appl
10	111	4.1	1036	US-09-536-224-4	Sequence 4, Appl
11	106	3.9	833	US-08-484-956-85	Sequence 85, Appl
12	106	3.9	833	US-08-757-653-85	Sequence 85, Appl
13	106	3.9	833	US-08-757-653-188	Sequence 188, Appl
14	106	3.9	833	US-08-823-516-69	Sequence 69, Appl
15	106	3.9	833	US-08-759-038-130	Sequence 130, Appl
16	106	3.9	833	US-08-758-314-130	Sequence 130, Appl
17	106	3.9	833	US-08-520-946-85	Sequence 85, Appl
18	105	3.9	833	US-08-757-653-168	Sequence 168, Appl
19	105	3.9	833	US-08-757-653-100	Sequence 100, Appl
20	105	3.9	833	US-08-823-516-66	Sequence 66, Appl
21	105	3.9	833	US-08-823-516-71	Sequence 71, Appl
22	105	3.9	833	US-08-759-038-107	Sequence 107, Appl
23	105	3.9	833	US-08-759-038-132	Sequence 132, Appl
24	105	3.9	833	US-08-758-314-107	Sequence 107, Appl
25	105	3.9	833	US-08-758-314-132	Sequence 132, Appl
26	101	3.7	812	US-08-446-794A-2	Sequence 2, Appl
27	101	3.7	812	US-08-750-007-3	Sequence 3, Appl

28	101	3.7	812	2	US-08-945-024-2	Sequence 2, Appl
29	100	3.7	553	4	US-09-096-399-2	Sequence 2, Appl
30	100	3.7	554	1	US-08-021-623C-6	Sequence 6, Appl
31	100	3.7	559	4	US-09-096-399-4	Sequence 4, Appl
32	100	3.7	832	1	US-07-977-434-2	Sequence 2, Appl
33	100	3.7	832	1	US-08-156-020-2	Sequence 2, Appl
34	100	3.7	832	1	US-08-156-020-4	Sequence 4, Appl
35	100	3.7	832	1	US-08-156-020-6	Sequence 6, Appl
36	100	3.7	832	1	US-08-156-020-8	Sequence 8, Appl
37	100	3.7	832	1	US-08-156-020-10	Sequence 10, Appl
38	100	3.7	832	1	US-08-073-384C-4	Sequence 4, Appl
39	100	3.7	832	1	US-08-254-359A-4	Sequence 4, Appl
40	100	3.7	832	1	US-08-483-043-4	Sequence 4, Appl
41	100	3.7	832	1	US-08-483-819-2	Sequence 2, Appl
42	100	3.7	832	1	US-08-481-238-4	Sequence 4, Appl
43	100	3.7	832	2	US-08-471-068B-4	Sequence 4, Appl
44	100	3.7	832	2	US-08-484-956-4	Sequence 4, Appl
45	100	3.7	832	2	US-08-757-653-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1	US-09-134-001C-4273	Sequence 4273, Application US/09134001C
Patent No. 6380370		
GENERAL INFORMATION:		
APPLICANT: Lynn Doucette-Stamm et al		
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC		
FILE REFERENCE: GTC-007		
CURRENT APPLICATION NUMBER: US/09/134,001C		
PRIOR FILING DATE: 1998-08-13		
PRIOR FILING DATE: 1997-11-08		
PRIOR APPLICATION NUMBER: US 60/064,964		
PRIOR FILING DATE: 1997-08-14		
NUMBER OF SEQ ID NOS: 5674		
SEQ ID NO 4273		
LENGTH: 495		
TYPE: PRT		
ORGANISM: Staphylococcus epidermidis		
US-09-134-001C-4273		
Query Match	10.9%; Score 295; DB 4; Length 495;	
Best Local Similarity	23.7%; Pred. No. 4.5e-21;	
Matches 103; Conservative 66; Mismatches 148; Indels 118; Gaps 14;		
QY 74 SDIYLLNDPHEAFARAALIESAHSLSLDQYYTMRNDISGRLLFNLYLAERGVRRL 133		
DB 132 NDLITTYDGNQKRDLDLNRHMOYTHIOYIIHSDNLKOLHLEKKAEGIEYKML 191		
QY 134 LDDNTRGLDLDLALDSHPNIEVRLNFPVLRKRRALGYLD--PP-----RLNR 183		
DB 192 YDMGSRDL-----RKDKLKKFRQGGHAESPSPKPLRLINLRMNR 233		
QY 184 MHKSTADNRATITLGRNIGDEYFVGEPTVA---DDILATGSVYGVSDFDRYNA 240		
DB 234 NHRKIIVDTIGTVGGFVNGDEY--IGSKKGYWRDHLRIKGAVALNQLRFLDWN 291		
QY 241 SHRAH-NATRIIR-----SGNIGKGLAIGNDETSRHALLRYRETVQPSLYOKIQ 292		
DB 292 SOSTRNLVYESYFPDVSOGTIGIOLASSGDE----- 326		
QY 293 GRIDWQVOTRLISDDPANGLDNRKRPPIAGLQDALQPEKSVYLVSPYPTKSGTD 352		
DB 327 ---DWEQIK-----YGLK-MISSAKSIVYQSPYFIPDQAFCD 361		
QY 353 ALAKLVQDGDYVTLNSLGATQVAHVSGYVYKRRPLKAGIKLYELQPNHVPATKDK 412		
DB 362 STKIALGVDVNIWPNKR--DHPEVYATLKNVASLLEAGVNVYHY-----DN 409		





STREET: 3031 Research Drive  
CITY: Richmond  
STATE: California  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/968,752B  
FILING DATE: 13-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/025,469  
FILING DATE: 05-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Giotta, Gregory J.  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX2004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-222-9700  
TELEFAX: 510-222-9758  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 932 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-968-752B-6

Query Match 4.8%; Score 130.5; DB 3; Length 932;  
Best Local Similarity 18.5%; Pred. No. 0.0006;  
Matches 104; Conservative 85; Mismatches 181; Indels 193; Gaps 25;

QY 86 FAARALIESAHS-----LDLOYI-----WRNDISGRLLFNLYLAERGAV 130  
DB 345 FAAYADALIRAOEELFTIDMWLSPELYLKRAHSDWRDI-----MLKRAEGVAV 397  
QY 131 RLDDNNTRGIDDLALLDS-----HPNIEVRLNPFVLRKRALGYLDFPRL 180  
DB 398 SILTFKE-----VELALGINSYSKRTMLLHPNIKV-MRHPDLVTLW-----439  
QY 181 NRRMINKSTADNRATILGGRNIG-----DEYFK---VGEDVFPADLLIATGS---VVG 229  
DB 440 --AHHEKLLVDDVVAFLGGLDLAFGRMDDVQYRLTDLGDPSEPVHLQPTLGSDPAATP 497  
QY 230 EYSHD-----FDRYASHAHNATRIIRSGNIGKGLQALGVNDETS 270  
DB 498 DLSHQFVFLGKDYSLNLTWKWVQDRPREDIDRETTFRMWRDVGVAHVGAARD-LA 556  
QY 271 RHALLRYETVE-----OSPLYOKI-----OTGRIDWOSVOTRLISDDPAKGLDRDRK 319  
DB 557 RHFIQRMWNTKTKKARYKPLPYLLPKSTSTANNLPFMIPEGQCATVQLSVDRWS--614  
QY 320 PPIARLQDA-----LKQPEKSYLVSPYFVPTKSTDLAKIVDDGID-----363  
DB 615 ---ACTLNSILINAYLHTIRESQHFLEYENOFIFISCSDRFVLNKGVEIDVRIKAMHQ 671  
QY 364 -----VTVLT-----NSLOATDVAAVHSGY-----V 384  
DB 672 GOCFVYVLLPLLPFGEEDISTGSGNSIO---ALHFTYRFLKRGESHILHRLKAAMGT 727  
QY 385 KYRKPLAKIGIKLYELQPNHAVPARKDKGLTSSVTSIAKFFIVDGKRIIFGSEFLDR 444  
DB 728 AMRDVNSICGLRTHGELGHP-----SELIYHSMKLLINDRRIIIGSANINIR 777  
QY 445 S--ARLNTMGVYIESPKIAEQMERTLADTSPRYAVRYTLDRNR-----LQ 489  
DB 778 SILGRDESLAILKD---TEMEPSLMGVYQAGRFALSLRGCFSVILGANTWPDLD 833  
QY 490 WHDPATRTKTYREPEAKIMKRIA 512

DB 834 LRDPVCDDF-----QLMOETA 850

RESULT 7  
US-09-536-224-6  
Sequence 6, Application US/09536224  
Patent No. 6379665  
GENERAL INFORMATION:  
APPLICANT: Frohman, Michael A.  
TITLE OF INVENTION: No. 6379665el Phospholipase D polypeptide and DNA sequences  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: ONYX Pharmaceuticals, Inc.  
STREET: 3031 Research Drive  
CITY: Richmond  
STATE: California  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,224  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/968,752  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Giotta, Gregory J.  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX2004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-222-9700  
TELEFAX: 510-222-9758  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 932 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-536-224-6

Query Match 4.8%; Score 130.5; DB 4; Length 932;  
Best Local Similarity 18.5%; Pred. No. 0.0006;  
Matches 104; Conservative 85; Mismatches 181; Indels 193; Gaps 25;

QY 86 FAARALIESAHS-----LDLOYI-----WRNDISGRLLFNLYLAERGAV 130  
DB 345 FAAYADALIRAOEELFTIDMWLSPELYLKRAHSDWRDI-----MLKRAEGVAV 397  
QY 131 RLDDNNTRGIDDLALLDS-----HPNIEVRLNPFVLRKRALGYLDFPRL 180  
DB 398 SILTFKE-----VELALGINSYSKRTMLLHPNIKV-MRHPDLVTLW-----439  
QY 181 NRRMINKSTADNRATILGGRNIG-----DEYFK---VGEDVFPADLLIATGS---VVG 229  
DB 440 --AHHEKLLVDDVVAFLGGLDLAFGRMDDVQYRLTDLGDPSEPVHLQPTLGSDPAATP 497  
QY 230 EYSHD-----FDRYASHAHNATRIIRSGNIGKGLQALGVNDETS 270  
DB 498 DLSHQFVFLGKDYSLNLTWKWVQDRPREDIDRETTFRMWRDVGVAHVGAARD-LA 556  
QY 271 RHALLRYETVE-----OSPLYOKI-----OTGRIDWOSVOTRLISDDPAKGLDRDRK 319  
DB 557 RHFIQRMWNTKTKKARYKPLPYLLPKSTSTANNLPFMIPEGQCATVQLSVDRWS--614  
QY 320 PPIARLQDA-----LKQPEKSYLVSPYFVPTKSTDLAKIVDDGID-----363

Db 615 ---AGLENSILNAYLHTTRESOHFLYIENOFISCSGDRVTLNKVGDDEIVDRILKAHQ 671  
QY 364 ---VTVLT-----NSLOATVNAVHSGY-----V 384  
Db 672 GOCFRYLLPLLPFGEGDISTGGNSIQ-----ALHFTYRLCRGESHILRLKAMGT 727  
QY 385 KYRKPILKAGIKLYELOPHNAVPATKDKLTGSSVTSIAKTFIVDGKRLFISFNIDPR 444  
Db 728 ARWDYMSIGLRTHGLGHPIT-----SELYIHSKMLADDRVYIGOSANINDR 777  
QY 445 S---ARLNTMGVYIESPKIAEQMERTLADTSPYAYRVTLDRHRN-----IQ 489  
Db 778 SLIGKDSSELALILIKD-----TEMEPSIMDGVEYQAGFALSILGRCFSVILGANTWPDLD 833  
QY 490 WHDPATRKTYPNPEAKIMKRIA 512  
Db 834 LRDPCVCDFF-----QLWQETA 850

RESULT 8  
US-09-107-149-2

; Sequence 2, Application US/09107149  
; Patent No. 6274363  
; GENERAL INFORMATION:  
; APPLICANT: Leung, David W.  
; APPLICANT: Tompkins, Christopher K.  
; TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D  
; FILE REFERENCE: 077319/014  
; CURRENT APPLICATION NUMBER: US/09/107,149  
; CURRENT FILING DATE: 1998-06-30  
; EARLIER APPLICATION NUMBER: 08/768,147  
; EARLIER FILING DATE: 1996-12-17  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 933  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-107-149-2

Query Match 4.5%; Score 121.5; DB 4; Length 933;

Best Local Similarity 17.8%; Pred. No. 0.0049;  
Matches 115; Conservative 90; Mismatches 194; Indels 247; Gaps 30;

QY 30 LKSCSS-----WLPLEERTSRHFNRSKPYRLNIIQI--RH---TPHTNGLSDIYLL 79  
Db 289 LLIKCSYQARWMOETELAOG-----PER--DFQLHRHDSYAPPRGTLARWY 339  
QY 80 NDPEHFAARALIESAESHLDQYI-----WRNDISGRLLFNLYLAEE 125  
Db 340 NGAGYAAVADILRAQEEIFITDMLSEVYLKRAHSDMDRLDI-----MLRKAE 392  
QY 126 RGYRVALLDDNNTRGDLDLLALDS-----HPNIEVLNPFYVLRKRALGYLT 175  
Db 393 EGYRVSILLFKE-----VELALGINSYSKALMLLHPNIKY--MRHDQYTW-- 439  
QY 176 DPFRLNRHNSFTADNRATILIG-----RNIGD----- 205  
Db 440 -----AHHEKLLVVDQYAVFAGLDLAYGRWDDLHYRLTDLGDSSESASOPPIPRPD 492  
QY 206 -----EYFKVGED-----TVFADLDLIATGSVY 228  
Db 493 SPATPDLSHNOFPWLGKQVSNLITNDWQLDREFEDFIDRETPRPMRWYGVYVAGLRA 552  
QY 229 GEVSHDFRYWASHAHNATRIIRSGNIGKGLALGYNDETSHALLRYREIVEGSPLIQ 288  
Db 553 RLARHFHFORW-----NFTKTTKA-----KYKTPYTYLLPKSTSTANQLPF-- 594  
QY 289 KIQTGRIQMSQVOTRLISDPAKGLDRDRKRPPIAGRLDA-----LKQEKSVYLV 340  
Db 595 TLPGGOC--TIVQV-----LKSVDKWS-----AGTLENSILNAYLHTTRESOHFLYIE 640

QY 341 SPYFVPTKSGTDALAKLVODGID-----VTVLTNSIQATD-----V 376  
Db 641 NOFFISCSGDRVTLNKVGEIYDRILKAHQGCYRYVLLPLLPFGEGDISTGGKSIQ 700  
QY 377 AAVHSGY-----VKYRKPILKAGIKLYELOPHNAVPATKDKLTGS 417  
Db 701 ALHFTYRLCRGEYSILRLKAMGTAMRDYISIGLRTHGLGHPV-----S 750  
QY 418 SVTSIAKTFIVDGKRLFISFNIDPRS--ARLNTMGVIE-----SPKI--AEQMERT 468  
Db 751 ELIYHSKVLADDRVYIIDSANINDRSLGKROSELAVILETETEPSIMNGAEYQAGR 810  
QY 469 LADTSPYAYRVTLDRHRN--LQWHPATRKTYPNPEAKIMKRIA 512  
Db 811 FALSIRKHCFSVILGANTRPDLDRPICDDFF-----QLWQDMA 850

RESULT 9  
US-08-968-752B-4

; Sequence 4, Application US/08968752B  
; Patent No. 6043073  
; GENERAL INFORMATION:  
; APPLICANT: Frohman, Michael A.  
; APPLICANT: Morris, Andrew  
; TITLE OF INVENTION: No. 6043073el Phospholipase D polypeptide and  
; TITLE OF INVENTION: DNA Sequences  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ONYX Pharmaceuticals, Inc.  
; STREET: 3031 Research Drive  
; CITY: Richmond  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94806  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/968,752B  
; FILING DATE: 13-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/025,469  
; FILING DATE: 05-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Giotta, Gregory J.  
; REGISTRATION NUMBER: 32,028  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 510-222-9700  
; TELEFAX: 510-222-9758  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1036 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-968-752B-4

Query Match 4.1%; Score 111; DB 3; Length 1036;

Best Local Similarity 21.5%; Pred. No. 0.068;  
Matches 128; Conservative 82; Mismatches 181; Indels 204; Gaps 38;

QY 60 NIQIHRPTHTNGLSDIYL-----LNDPEHFAARALIESAESHLDQYIWRND- 110  
Db 444 NIKVMHHPHVS--STVYILMAHHEKILVIIDQSVAFVG-----GIDLAYGRWDDNE 491  
QY 111 -----ISGULFNLVYLAERGVRYRLDDNN-----TRGLDLLLLAD 150  
Db 492 HRLTGVGSKRVYSGPSGLSLPPAAMESMELR--LKDKNPEVQNLPIQKSIDVDVSKLK 549

```

QY 151 SHNIEVRLFNPPVLLKRWALGYLTDFPRLNRMHNSFTADNRATILG----- 199
DB 550 GIG--KPRKFSKSLYKOLHRLHLDADSI--SIDSTNTGSIKSLQGTGELHGETRFW 606
QY 200 -GRNIGDEXEK--VGEDTVFADL-----DILAT--GSVVEGESHDEDRYASH 242
DB 607 HGKDYCNFVKDWYQDLKDFADRIDRYSTPRMWHDIASVHGKARDVARRHFIQRM--- 663
QY 243 SAHNATRIIRSGNIGKGLALGY-----NDETSRHALLRYRETEQSP--LYOKIOTGR- 294
DB 664 ---NFTKIMKS-----KYRSLSYPELLPKSQTTAHE-LRY-----QVPSGVHANVOLLRS 709
QY 295 -IDW-----OSVOTRLI-----SDPPAGLDRDRKRPPIAGR 325
DB 710 AADWMSGICYHEESHIAAIVHIENSRIHYIYENOFFISCADK---VFNKIGDAIAOR 766
QY 326 LODALKOPEK-SVYLVSPEYFP-----TKSGTDLAKLVODGIDVTYL--TNSLOATDVA 377
DB 767 ILKAHRENGKYRYVYIP--LLPGFEGDISTGGNALQAIMHFNRYRMCGENSIIGQLKA 825
QY 378 AVHSGYVKRKRLKAGIKLYELQPNHNAVPAKDKGLTSSVTSL---HAKFTIYDGKRI 434
DB 826 ELGNOMINY--ISFCGLR-----THA-----ELGSLVTELYVHSHKLIADNDTV 869
QY 435 FIGSFNLDPRS--ARLNTMGVIES-----PRIAQMERTLADTSPEY----- 476
DB 870 IIGSANINDRSMLGKDSMAVIVDTEVPSYMD-----GKEYQAGRFAGRLQ 920
QY 477 AYRVTL---DHRNLQWMDPATRTKTYPNPEPAKLMKRIA-----KILSLP 520
DB 921 CRRVYLGYLDPESEDIQ--DPVSDKFF-----KEYWVSTAARNATLYDKVFRCLP 968

```

```

RESULT 10
US-09-536-224-4
; Sequence 4, Application US/09536224
; Patent No. 6379665
; GENERAL INFORMATION:
; APPLICANT: Frohman, Michael A.
; TITLE OF INVENTION: No. 6379665el Phospholipase D Polypeptide and
; TITLE OF INVENTION: DNA Sequences
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONYX Pharmaceuticals, Inc.
; STREET: 3031 Research Drive
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94806
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,224
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/968,752
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Giotta, Gregory J.
; REGISTRATION NUMBER: 32,028
; REFERENCE/DOCKET NUMBER: ONYX2004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-222-9700
; TELEFAX: 510-222-9758
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1036 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-536-224-4

```

```

Query Match 4.1%; Score 111; DB 4; Length 1036;
Best Local Similarity 21.5%; Pred. No. 0.066;
Matches 128; Conservative 82; Mismatches 181; Indels 204; Gaps 38;

```

```

QY 60 NILQIRPPTNGLSDIYL-----LNDPHEAFARAALIESAHSIDLQYIRND- 110
DB 444 NIKVMHPPHVS--STVYLMAHHEKLYIIDQSAFVG-----GIDLAVGRDDNE 491
QY 111 -----ISGRLENLYLAAERGVRRLLIDNN-----TRGLDILLALD 150
DB 492 HRLTDVSVKRYVSGPSLSPAPAMESMSLR--LKDNEPQNPILKSIDVDVSKLK 549
QY 151 SHNIEVRLFNPPVLLKRWALGYLTDFPRLNRMHNSFTADNRATILG----- 199
DB 550 GIG--KPRKFSKSLYKOLHRLHLDADSI--SIDSTNTGSIKSLQGTGELHGETRFW 606
QY 200 -GRNIGDEXEK--VGEDTVFADL-----DILAT--GSVVEGESHDEDRYASH 242
DB 607 HGKDYCNFVKDWYQDLKDFADRIDRYSTPRMWHDIASVHGKARDVARRHFIQRM--- 663
QY 243 SAHNATRIIRSGNIGKGLALGY-----NDETSRHALLRYRETEQSP--LYOKIOTGR- 294
DB 664 ---NFTKIMKS-----KYRSLSYPELLPKSQTTAHE-LRY-----QVPSGVHANVOLLRS 709
QY 295 -IDW-----OSVOTRLI-----SDPPAGLDRDRKRPPIAGR 325
DB 710 AADWMSGICYHEESHIAAIVHIENSRIHYIYENOFFISCADK---VFNKIGDAIAOR 766
QY 326 LODALKOPEK-SVYLVSPEYFP-----TKSGTDLAKLVODGIDVTYL--TNSLOATDVA 377
DB 767 ILKAHRENGKYRYVYIP--LLPGFEGDISTGGNALQAIMHFNRYRMCGENSIIGQLKA 825
QY 378 AVHSGYVKRKRLKAGIKLYELQPNHNAVPAKDKGLTSSVTSL---HAKFTIYDGKRI 434
DB 826 ELGNOMINY--ISFCGLR-----THA-----ELGSLVTELYVHSHKLIADNDTV 869
QY 435 FIGSFNLDPRS--ARLNTMGVIES-----PRIAQMERTLADTSPEY----- 476
DB 870 IIGSANINDRSMLGKDSMAVIVDTEVPSYMD-----GKEYQAGRFAGRLQ 920
QY 477 AYRVTL---DHRNLQWMDPATRTKTYPNPEPAKLMKRIA-----KILSLP 520
DB 921 CRRVYLGYLDPESEDIQ--DPVSDKFF-----KEYWVSTAARNATLYDKVFRCLP 968

```

```

RESULT 11
US-08-484-956-85
; Sequence 85, Application US/08484956
; Patent No. 5843654
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; APPLICANT: OLDENBURG, MARY C.
; TITLE OF INVENTION: DETECTION OF p53 MUTATIONS
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAYESTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25

```



```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,956
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/402,601
FILING DATE: 09-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,164
FILING DATE: 09-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/254,359
FILING DATE: 06-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 04-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, J, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01801
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-956-85

```

```

Query Match          3.9%; Score 106; DB 2; Length 833;
Best local Similarity 21.8%; Pred. No. 0.15;
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

```

```

QY 118 NLVYLAERGVRY-----RLLDNNTRGL---DDLALDS---HPNIEVRLNPF 163
DB 321 DLALAAAGRGVHRAPEPKALDLKEARGLAKDSLVALRGGLPGDDPML----- 376
QY 164 VLKWRALGLYLD---FPLNRMHKST--ADNRAT-----LGGRNIGDE----- 206
DB 377 -----LAYLDPSNTTPGVARRYGGEWTEEGERAALSERLPANLWGRLEGEERLLW 429
QY 207 -YFKVGE--DYVADLDILATG-----SYGGEVS-----HDFDRYW 239
DB 430 LYREVERPLSAVLAHME--ATGVRLDVAYLRLALSLEVAGEIARLEAEVFRLAGHPFN-- 484
QY 240 ASHSAHNATRIINSNGIGLQALGYNDETSRIA-----LLRYRETV 281
DB 485 -LNSROLEVLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIYEKILQYRELT 540
QY 282 EQ-----SPLYOKI--OTGRIDMOSVOT-----RLISDDPAKGLDRRRKPIAGRLQDA 329
DB 541 KLSKSTYIDPLDLIHPRTRGLHTRFNOTATATGRLSSDP--NLQNIPTVPTLQGRIRRA 598
QY 330 LKREKSVILVSPY-----FVPTKSGTDALAKLYODGIDIVYVLNS-----LQATDVA 377
DB 599 FLAEEMWLVALDYSOIELKVLHLSGDENLINVFOGRIHTETJASWMEGVPREAVD-- 656
QY 378 AVHSGVYKVRPLKA-----GIKLYELQPNHAPV----- 407
DB 657 -----PLMRAAKTINFGVLYKMSAHLISOELALPYEAOAFIERIYQSPFKVR 705
QY 408 ----ATKDGGLGSSVTSLSHAKTFIYDKRIFIGSENLDRSAPLNTGMCVITESP--K 460
DB 706 AWIEKLELEBRRGVYETLL-----GRRRYVDELARVKSVEAER--MAFMMPVPGT 757
QY 461 IAAOMERTLADTSP---EYAVRYTLDRHNLQMHDPATRTKTYNEPEAKLMKRIAAITLS 517

```

```

DB 758 AADLMKIAWVKLPRLREEMGARMILQVHDELVLEAPKERA-----EAVARLAKEVMEGVYP 813
QY 518 L-LPIE 522
DB 814 LAVPPL 819

```

```

RESULT 12
US-08-757-653-85
Sequence 85, Application US/08/757653
Patent No. 5843669
GENERAL INFORMATION:
APPLICANT: Kaiser, Michael W.
APPLICANT: Lyamichiev, Victor I.
TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESS: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/757,653
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02565
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 397-8338
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 85:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-757-653-85

```

```

Query Match          3.9%; Score 106; DB 2; Length 833;
Best local Similarity 21.8%; Pred. No. 0.15;
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

```

```

QY 118 NLVYLAERGVRY-----RLLDNNTRGL---DDLALDS---HPNIEVRLNPF 163
DB 321 DLALAAAGRGVHRAPEPKALDLKEARGLAKDSLVALRGGLPGDDPML----- 376
QY 164 VLKWRALGLYLD---FPLNRMHKST--ADNRAT-----LGGRNIGDE----- 206
DB 377 -----LAYLDPSNTTPGVARRYGGEWTEEGERAALSERLPANLWGRLEGEERLLW 429
QY 207 -YFKVGE--DYVADLDILATG-----SYGGEVS-----HDFDRYW 239
DB 430 LYREVERPLSAVLAHME--ATGVRLDVAYLRLALSLEVAGEIARLEAEVFRLAGHPFN-- 484
QY 240 ASHSAHNATRIINSNGIGLQALGYNDETSRIA-----LLRYRETV 281
DB 485 -LNSROLEVLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIYEKILQYRELT 540
QY 282 EQ-----SPLYOKI--OTGRIDMOSVOT-----RLISDDPAKGLDRRRKPIAGRLQDA 329
DB 541 KLSKSTYIDPLDLIHPRTRGLHTRFNOTATATGRLSSDP--NLQNIPTVPTLQGRIRRA 598

```

QY 330 LKQPKSVYLVSPY-----FVPTKSGTALAKLVODGIDVTLVNS-----LOATDVA 377  
 DB 599 FIAEEGMLVALADYSOIELRLVLAHLSGDENLIVFQEGRD IHETASMGVPREAVD-- 656  
 QY 378 AVHSGYVKKRKLKA-----GIKLYELOPNHVP----- 407  
 DB 657 -----PLMRAAKTINFGVLYGMSAHLRISOELAIPEBAQAFIERYPQSPKVR 705  
 QY 408 ----ATKDKGLTSSVTSLSHAKTFIVDGKRIPIGSENLDPKRSARLNTMGVIESP---K 460  
 DB 706 AMIEKTELEGGRRRGVYETLF-----GRRRYPDLEARKSVREAAR-MAFNMPVQGT 757  
 QY 461 IAEQMERTLADTSP---EYAYRVTLDNRHRLQWHDPAKTRKTYNEPEAKIMKRIAKILS 517  
 DB 758 ADLMKLAVALKFLPRLIEBGARMLLOVHDELVEAKRERA---EVAVARLAKVEMGVYPR 813  
 QY 518 L-LPIE 522  
 DB 814 LAVPLE 819

## RESULT 13

US-08-757-653-188  
 ; Sequence 188, Application US/08757653  
 ; Patent No. 5843669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kaiser, Michael W.  
 ; APPLICANT: Lyamichev, Victor I.  
 ; APPLICANT: Lyamichev, Natasha  
 ; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using  
 ; TITLE OF INVENTION: Thermostable FEN-1 Endonucleases  
 ; NUMBER OF SEQUENCES: 190  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medien & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States Of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/757,653  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ingolia, Diane E.  
 ; REGISTRATION NUMBER: 40,027  
 ; REFERENCE/DOCKET NUMBER: FORS-02565  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 705-8410  
 ; TELEFAX: (415) 397-8338  
 ; INFORMATION FOR SEQ ID NO: 188:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 833 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-757-653-188

Query Match 3.9%; Score 106; DB 2; Length 833;  
 Best Local Similarity 21.8%; Pred. No. 0.15;  
 Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

QY 118 NLVYLAENGAVY-----RLILDNNTGRL---DDLALDS---HPNIEVRLNPF 163  
 DB 321 DLLAALAANGGVHRAPEYKALRDLKEARGLLADLSVALREGILGDPDPMI----- 376  
 QY 164 VLRKRALGYLND----FPRLNRBMHNSFT--ADNRATI-----LGRNIGDE----- 206

DB 377 -----LAVLLDPSNTTPEGVARRYGGEWTEBAGERRAALSERLIFANIMORLEGEBRLM 429  
 QY 207 -YFKYGE--DTVRADIDILATG-----GVGVS-----HDDRWA 239  
 DB 430 LYREVERPLSAVLAHNE--ATGVRLDVAIVRLSLSEVAGEIARLEAEVRLAGHPN--- 484  
 QY 240 ASHSANNAIRIIRSGNIGKQALQALYNDETSRA-----LLRYRETV 281  
 DB 485 -LNSRQLEEVLD---ELGLPAIGKTEKRGKSTSAVLEALREAHPIYEKILQYBELT 540  
 QY 282 EQ-----SPLQKI--QTGRIDMOSYQT-----RLISDPAKLDNRKRPPIAGLQDA 329  
 DB 541 KIKSTYIDPLDILHPTGRLHTRFNOTATANGRLSSDP--NLQNIPTVPLGGRIRRA 598  
 QY 330 LKQPKSVYLVSPY-----FVPTKSGTALAKLVODGIDVTLVNS-----LOATDVA 377  
 DB 599 FIAEEGMLVALADYSOIELRLVLAHLSGDENLIVFQEGRD IHETASMGVPREAVD-- 656  
 QY 378 AVHSGYVKKRKLKA-----GIKLYELOPNHVP----- 407  
 DB 657 -----PLMRAAKTINFGVLYGMSAHLRISOELAIPEBAQAFIERYPQSPKVR 705  
 QY 408 ----ATKDKGLTSSVTSLSHAKTFIVDGKRIPIGSENLDPKRSARLNTMGVIESP---K 460  
 DB 706 AMIEKTELEGGRRRGVYETLF-----GRRRYPDLEARKSVREAAR-MAFNMPVQGT 757  
 QY 461 IAEQMERTLADTSP---EYAYRVTLDNRHRLQWHDPAKTRKTYNEPEAKIMKRIAKILS 517  
 DB 758 ADLMKLAVALKFLPRLIEBGARMLLOVHDELVEAKRERA---EVAVARLAKVEMGVYPR 813  
 QY 518 L-LPIE 522  
 DB 814 LAVPLE 819

## RESULT 14

US-08-823-516-69  
 ; Sequence 69, Application US/08823516  
 ; Patent No. 5994069  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hall, Jeff G.  
 ; APPLICANT: Lyamichev, Victor I.  
 ; APPLICANT: Mast, Andrea L.  
 ; APPLICANT: Brow, Mary Ann D.  
 ; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
 ; TITLE OF INVENTION: Sequential Invasive Cleavages  
 ; NUMBER OF SEQUENCES: 163  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Medien & Carroll, LLP  
 ; STREET: 220 Montgomery Street, Suite 2200  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States Of America  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/823,516  
 ; FILING DATE: 24-MAR-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US97/01072  
 ; FILING DATE: 21-JAN-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/759,038  
 ; FILING DATE: 02-DEC-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/758,314  
 ; FILING DATE: 02-DEC-1996

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/756,386
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02736
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 69:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-823-516-69

```

```

Query Match          3.9%; Score 106; DB 2; Length 833;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

```

```

QY 118 NLVYLAERGAV-----RLLDNNRGL---DDLALDS---HPNIEVRLFNPF 163
DB 321 DLLALAAAGGVHRAPEPYKALRDKEARGLANDLSVALREGIGLPDDPMI---- 376
QY 164 VLKRRALGYLTD---FPRLNRRMHKSF---ADNRATI-----LGGRNIGDE--- 206
DB 377 -----LAYLLDPSNTPEGVARRYGGEWTEAGERALSERLRANLWGRLEGERLLM 429
QY 207 -YFKVGE--DIVFADDLIATG-----SVGEVS-----HDFDRYW 239
DB 430 LYREVERPLSAVLHME--ATGVRLDVAVALRALSLEVAGEIARLEAEVRLAGHPFN--- 484
QY 240 ASHSAHNATRIIRSGNIGKGLQALGYNDETSRA-----LTRYRETV 281
DB 485 -LNSRDQLERYLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIVEKILQIRELT 540
QY 282 EQ-----SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRDRKKPIAGRLQDA 329
DB 541 KIKSTYIDPLPLIHPRIQRGLHTRNQGTATGRSSSP--NLQNIPIRTLGGQIRIRA 598
QY 330 LKQPEKSYLVSPY-----FVPTKSGTDALAKLVQDGIQVTLTNS-----LQATDVA 377
DB 599 FTAEGMWLVALAYSOIELRLVLAHLSGDENLIRVFOEGRDITETASWMEGVPREAVD-- 656
QY 378 AVHSGVVKRRLKA-----GIKLEYLQPNHNVF----- 407
DB 657 -----PLMKRAKTIQFVLYGSAHRLSOELIPIEEAOAFIERIFQSEPKVR 705
QY 408 ---ATKDKGLTGSVTSLSHAKTFTVDGKRIFIGSFNLDPRSARLNTKGVYIESP---K 460
DB 706 AMIETLEGGRRRGVEYELF-----GRRRYVPLEARVKKVREAEER-MAFNMPVQGT 757
QY 461 IAEQMERLTADTSP---EYAYRVTLDRHNRLQWMDPATRKTYVNEPEAKLMKRIAKTILS 517
DB 758 AADLKLAMVKKLFPRLEEGARMLLOVHDELVLAPKERA---BAVAVRLAKEVMEGVYP 813
QY 518 L-LPIE 522
DB 814 LAVPLE 819

```

```

RESULT 15
US-08-759-038-130
Sequence 130 Application US/08759038
Patent No. 6090343

```

```

; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; APPLICANT: Dahlberg, James E.
; TITLE OF INVENTION: Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 134
; CORRESPONDENCE ADDRESS:
; ADDRESS: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,038
; FILING DATE: 02-DEC-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/
; FILING DATE: 29-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-759-038-130

```

```

Query Match          3.9%; Score 106; DB 3; Length 833;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

```

```

QY 118 NLVYLAERGAV-----RLLDNNRGL---DDLALDS---HPNIEVRLFNPF 163
DB 321 DLLALAAAGGVHRAPEPYKALRDKEARGLANDLSVALREGIGLPDDPMI---- 376
QY 164 VLKRRALGYLTD---FPRLNRRMHKSF---ADNRATI-----LGGRNIGDE--- 206
DB 377 -----LAYLLDPSNTPEGVARRYGGEWTEAGERALSERLRANLWGRLEGERLLM 429
QY 207 -YFKVGE--DIVFADDLIATG-----SVGEVS-----HDFDRYW 239
DB 430 LYREVERPLSAVLHME--ATGVRLDVAVALRALSLEVAGEIARLEAEVRLAGHPFN--- 484
QY 240 ASHSAHNATRIIRSGNIGKGLQALGYNDETSRA-----LTRYRETV 281
DB 485 -LNSRDQLERYLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIVEKILQIRELT 540
QY 282 EQ-----SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRDRKKPIAGRLQDA 329
DB 541 KIKSTYIDPLPLIHPRIQRGLHTRNQGTATGRSSSP--NLQNIPIRTLGGQIRIRA 598
QY 330 LKQPEKSYLVSPY-----FVPTKSGTDALAKLVQDGIQVTLTNS-----LQATDVA 377

```

Db 539 FIAEGWLLVALAVSOLRLVLAHLSGDENLIRVQEGDITETASWFGVPREAVD-- 656  
QY 378 AVHSGYVYRRPLKA-----GIKYLEOPNHAVP----- 407  
Db 657 -----PLMRRAKTIHFGVLYGMSAHRLSQELAIPEEQAFIERFQSFPKVR 705  
QY 408 ---ATKDKGLTSSVTSILHAKFTVDGKRIFIGSFNDPRSARLNTMGVIESP---K 460  
Db 706 AMIEKTELEGGRRRGVETLF-----GRRRYVPLDLARVKSVRBAER-MAFNMPVOGT 757  
QY 461 IAEQWERTLADTSP---EYAVYVTLDRHNRLQWHPATRKTYPNPEAKLMKRIAKILS 517  
Db 758 AADLMKLAHVLFPRLEEMGARMLQVHDELVLPAKERA---EAVARLAKVEWEGYVP 813  
QY 518 L-LPIE 522  
Db 814 LAVPLE 819  
RESULT 16  
US-08-758-314-130  
; Sequence 130, Application US/08758314  
; Patent No. 6090606  
; GENERAL INFORMATION:  
; APPLICANT: Kaiser, Michael W.  
; APPLICANT: Lyamichev, Victor I.  
; APPLICANT: Lyamichev, Natasha  
; TITLE OF INVENTION: Improved Cleavage Agents  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medien & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States Of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/758,314  
; FILING DATE: 02-DEC-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/  
; FILING DATE: 29-NOV-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/682,853  
; FILING DATE: 12-JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,491  
; FILING DATE: 24-JAN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: FORS-02575  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEO ID NO: 130:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 833 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-758-314-130  
Query Match 3.9%; Score 106; DB 3; Length 833;  
Best Local Similarity 21.8%; Pred. No. 0.15;  
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

QY 118 NLVYLAERGVY-----RLLDNNNTGVL---DDLALDS---HPNIEVLEPNF 163  
Db 321 DLIALAARGGVHRAPEPYKALRDLKARGLAKDLSVALRBSGLPPODDPML----- 376  
QY 164 VLKRWALGYLTD---FRLNRHNSFT---ADNRAT-----LGNNDGE----- 206  
Db 377 ---LAVLDPSTNTPEGVARYGGEWTEEAGERAALSERLFANIMGRLEGEERLLM 429  
QY 207 -YFKVGE--DIVFEDLDLTLG-----SYVGEVS-----HDEFDYX 239  
Db 430 LYREVERPLSAVLAHE--ATGVRLDVAVYRALSLIEVAGETARLEAEVYRLAGHPFN--- 484  
QY 240 ASHSAHNAIRIIRSGNIGKGLQALGYNDSTRNA-----LTRYRETV 281  
Db 485 -LNSRDQLEEVLFD---ELGLPAIKTEKTKGRSTSAVALEALREAPHIVKILQYREL 540  
QY 282 EQ-----SPLYOKT--QGRIDMOSVQT-----RLSDPPAKGLDRDRKRPPIAGRLDA 329  
Db 541 RLKSTYIDPLDPLHPRTRGLHTRFNQATATAGRLSSDP--NLQNIPLVPTPLQRIIRA 598  
QY 330 LKOPEKSYVLVSPY-----FVPTKSGTDALAKLVQSDIDVTVLNS-----LQATDVA 377  
Db 599 FIAEGWLLVALAVSOLRLVLAHLSGDENLIRVQEGDITETASWFGVPREAVD-- 656  
QY 378 AVHSGYVYRRPLKA-----GIKYLEOPNHAVP----- 407  
Db 657 -----PLMRRAKTIHFGVLYGMSAHRLSQELAIPEEQAFIERFQSFPKVR 705  
QY 408 ---ATKDKGLTSSVTSILHAKFTVDGKRIFIGSFNDPRSARLNTMGVIESP---K 460  
Db 706 AMIEKTELEGGRRRGVETLF-----GRRRYVPLDLARVKSVRBAER-MAFNMPVOGT 757  
QY 461 IAEQWERTLADTSP---EYAVYVTLDRHNRLQWHPATRKTYPNPEAKLMKRIAKILS 517  
Db 758 AADLMKLAHVLFPRLEEMGARMLQVHDELVLPAKERA---EAVARLAKVEWEGYVP 813  
QY 518 L-LPIE 522  
Db 814 LAVPLE 819  
RESULT 17  
US-08-520-946-85  
; Sequence 85, Application US/08520946  
; Patent No. 6372424  
; GENERAL INFORMATION:  
; APPLICANT: BROW, MARY ANN D.  
; APPLICANT: LYAMICHEV, VICTOR I.  
; APPLICANT: OLIVE, DAVID M.  
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF  
; NUMBER OF SEQUENCES: 160  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MEDIEN & CARROLL  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/520,946  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CARROLL, PETER G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: FORS-01756

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 85:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-520-946-85

Query Match      3.9%; Score 106; DB 4; Length 833;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

QY 118 NLVYLAEGVRY-----RLLDNNTRGL---DDLALDS---HPNIEVRLFP 163
DB 321 DLLALAAAGRVHRAPEYKALRDLEKARGLAKDLSVALREGILPPGDPM 376
QY 164 VLKRNALGYLD---FRLNRHMKSF--ADNRATI-----LGRNIGDE---- 206
DB 377 -----LAYLDPSNTPEGVARRYGEMTEBAGERALSERLFLANLGRLEGERLLM 429
QY 207 -YFKVGE--DVFADLDILATG-----SVGEVS-----HDEDRYW 239
DB 430 LYREVERPLSAVLAHNE--ATGVRLDVAYLRLSLLEVAGELARLEAEVFLAGHPEN--- 484
QY 240 AASHAHNARRITRSNIGKGLQALGNDTSRHA-----LKRREIV 281
DB 485 -LNSRDQLERVLFD---ELGLPAIGKTEKRGKSTSAVLEALREAHPIVEKILQYREL 540
QY 282 EQ-----SPLYOKI--QGRIDMOSVOT-----RLISDPAGLDNRKRPPIAGRLQDA 329
DB 541 KIKSTYIDPLPLHRTGRHLTRFNQTATGRLLSSDP--NLQNIPTPLGQRIRRA 598
QY 330 LKQPEKSYLVSPY-----FVPTKSGTDALAKLYODGIDVTVLNS-----LQATDVA 377
DB 599 FLAEEMWLVALDYQIELRLVLAHLSGDENLIRVFEQGDHITETASWMEGVPREAVD-- 656
QY 378 AVHSGVYKRPDLKA-----GIKLYELQPNHNP----- 407
DB 657 -----PLMRAAKTINFGVLYGSAHRLSOELAPIPEEAQAFIERFYQSPKVR 705
QY 408 ---ATKDKGLTSSVTSIAHAKTFIVDGKRIFIGSFNLDPRSARLNTGVVIESP---K 460
DB 706 AMIEKTLIEGRRRGVETLF-----GRRRYVPDLKAVKSVREAER--MAFNMPVRGT 757
QY 461 IAEQERTIADTSP---EYAYRVTLDRHNRLOWHDPATRKTYPNPEAKLMKRIAKILIS 517
DB 758 AADLMKLAIVKLFPLEMGARMLQVHDELVLKPKERA---EAVARLAEVMEGVYP 813
QY 518 L-LPIE 522
DB 814 LAVPLE 819

RESULT 18
US-08-757-653-168
; Sequence 168, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamlichev, Victor I.
; APPLICANT: Lyamlichev, Natasha
; TITLE OF INVENTION: Cleavage Of Nucleic Acid Using
; TITLE OF INVENTION: Thermostable Fen-1 Endonucleases
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California

```

```

; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,653
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02565
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 168:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 833 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-757-653-168

Query Match      3.9%; Score 105; DB 2; Length 833;
Best Local Similarity 21.5%; Pred. No. 0.19;
Matches 116; Conservative 65; Mismatches 183; Indels 176; Gaps 27;

```

```

QY 118 NLVYLAEGVRY-----RLLDNNTRGL---DDLALDS---HPNIEVRLFP 163
DB 321 DLLALAAAGRVHRAPEYKALRDLEKARGLAKDLSVALREGILPPGDPM 376
QY 164 VLKRNALGYLD---FRLNRHMKSF--ADNRATI-----LGRNIGDE---- 206
DB 377 -----LAYLDPSNTPEGVARRYGEMTEBAGERALSERLFLANLGRLEGERLLM 429
QY 377 -----LAYLDPSNTPEGVARRYGEMTEBAGERALSERLFLANLGRLEGERLLM 429
QY 207 -YFKVGE-----DVFADLDILATGVSVEGSHDPRYAS-----HSAH 245
DB 430 LYREVERPLSAVLAHNEATGVRLDVAYLRLSLLEVAGELARLEAEVFLAGHPFNLSND 489
QY 430 LKQPEKSYLVSPY-----FVPTKSGTDALAKLYODGIDVTVLNS-----LQATDVA 377
DB 430 LKQPEKSYLVSPY-----FVPTKSGTDALAKLYODGIDVTVLNS-----LQATDVA 377
QY 246 NATRIIRSNIGKGLQALGNDTSRHA-----LKRREIV 281
DB 490 QLEERVLFD---ELGLPAIGKTEKRGKSTSAVLEALREAHPIVEKILQYREL 540
QY 284 -SPLYOKI--QGRIDMOSVOT-----RLISDPAGLDNRKRPPIAGRLQDA 329
DB 547 IDPLPLHRTGRHLTRFNQTATGRLLSSDP--NLQNIPTPLGQRIRRA 598
QY 336 SYLVSPY-----FVPTKSGTDALAKLYODGIDVTVLNS-----LQATDVA 377
DB 605 WLLVALDYQIELRLVLAHLSGDENLIRVFEQGDHITETASWMEGVPREAVD-- 656
QY 384 VKTRKPLKA-----GIKLYELQPNHNP----- 409
DB 657 -----PLMRAAKTINFGVLYGSAHRLSOELAPIPEEAQAFIERFYQSPKVR 705
QY 410 KDKGLTSSVTSIAHAKTFIVDGKRIFIGSFNLDPRSARLNTGVVIESP---K 460
DB 712 LEEGRRRGVETLF-----GRRRYVPDLKAVKSVREAER--MAFNMPVRGT 757
QY 467 RLADTSP---EYAYRVTLDRHNRLOWHDPATRKTYPNPEAKLMKRIAKILIS 517
DB 764 LAMVKLFPLEMGARMLQVHDELVLKPKERA---EAVARLAEVMEGVYP 819

RESULT 19
US-08-757-653-190
; Sequence 190, Application US/08757653
; Patent No. 5843669
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.

```

APPLICANT: Lyamichev, Victor I.  
APPLICANT: Lyamichev, Natasha  
TITLE OF INVENTION: Cleavage Of Nucleic Acid Using  
TITLE OF INVENTION: Thermostable FEN-1 Endonucleases  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,653  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02565  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 190:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-757-653-190

Query Match 3.9%; Score 105; DB 2; Length 833;  
Best Local Similarity 21.8%; Pred. No. 0.19; Indels 188; Gaps 30;  
Matches 119; Conservative 65; Mismatches 174;

QY 118 NLVYLAERGVV-----RLLDNNRTGL--DDLALDS---HPNIEVLEFNP 163  
DB 321 DLALAAARGVHRAPRYKALRLKEARGLAKDLVSLAREGLGPRGDDPL---- 376  
QY 164 VLKRRALGYLD-----EPRLNRKHNKSF--ADNRATI-----LGRRTGE---- 206  
DB 377 -----LAVLLDPSNTTPEGVARYGGEWTEAGERALSRLPANLWGREGELRLW 429  
QY 207 -YFKVGE--DTVFADLLIATG-----SVVGEVS-----HDEDRYW 239  
DB 430 LYREVERLSAVLAHME--ATGVRLDVAYLRLSLVAGELRLAEVFRLAGHFFN--- 484  
QY 240 ASHSANATRIIRSGNIGKGLQALQYNDETSRA-----LRLRYRTV 281  
DB 485 -LNSHDLERVLFD---ELGLPAIGKTEKTKGSTSAVLEALREAHPIVERKILQYREL 540  
QY 282 EQ-----SPVQKI--QGRIDMOSVQ-----RLISDDPAKGLDRKRPPIARLQDA 329  
DB 541 KIKSTYIDPLDLPKRLGRILHTRNGATATATGRILSSSDP--NLQNIPIVRYPLGGRIRRA 598  
QY 330 LKQEKSVYLVSPY-----FVPTKSGDALAKLVQDGIDVTLTNS-----LQATDVA 377  
DB 599 FIAEGRGLLVALVYSQILRLVLAHLSGDENLIRVQEGRDITHTASMMFGVPREAVD-- 656  
QY 378 AVHSGYVYKRRPLKA-----GIKLYELOPNHAPV----- 407  
DB 657 -----PLMRRAKTIINFGVLYGMSAHRISQELAPIPEEAQAFIERYFQSPKVR 705  
QY 408 -----ATKRGKLTGSSVYSLAKTFLVDCGRIFIGSFNLDPSNARLNTKGVYIESP---K 460  
DB 706 AMIEKTLTEEGRRRGYVETLF-----GRRRVVPLEARVKSVSREAAER--MAFNMPVOGT 757  
QY 461 IAEQWERTLADTSP---EYAVRVTLDRHNRLQWMDPATRKTYPNPEEAKLWKRIAKILIS 517

DB 758 AADLKKLAVKLFPRLEEMGARMLOVHDELVLEAPKERA-----EAVARLAEVMEGYVP 813  
QY 518 L-LPIE 522  
DB 814 LAVPIE 819

RESULT 20  
US-08-823-516-66  
Sequence 66; Application US/08823516  
Patent No. 5994069  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Mast, Andrea L.  
APPLICANT: Brow, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-823-516-66

Query Match 3.9%; Score 105; DB 2; Length 833;  
Best Local Similarity 21.5%; Pred. No. 0.19; Indels 183; Gaps 27;  
Matches 116; Conservative 65; Mismatches 183;



Query Match	3.9%	Score 105;	DB 3;	Length 833;
Best Local Similarity	21.8%;	Pred. No. 0.19;		
Matches 119; Conservative	65;	Mismatches 174;	Indels 188;	Gaps 30





```

STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,314
FILING DATE: 02-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02575
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-758-314-132

Query Match
Best Local Similarity 3.9%; Score 105; DB 3; Length 833;
Matches 119; Conservative 65; Mismatches 174; Indels 186; Gaps 30;

QY 118 NLVYLAERGVR-----RLDDNNTGRL--DGLLALDS---HENIEVRLNPF 163
DB 321 DILLALAARGVRHRAPEYKALRLDLKEARGLAKDLVLALEGGLPGDDPML---- 376
QY 164 VRKRNALGYLD---FRLNRHMNKSFT--ADNRATL-----LGRNIGDE---- 206
DB 377 -----LAVLLDPSNTTEGVARRRGWTEBAGERALSERLFWANLGRLEGEERLLM 429
QY 207 -YFKVGE--DTVEADDILATG-----SVGEVS-----HDDRYW 239
DB 430 LYREVRPLSAVLAHME--ATGYRLDVATLRALSLEVAGEIARLEAEVRLAGHPN--- 484
QY 240 ASHSAHNAATRIIRSGNIGKGLALGYNDETSRHA-----LLRYRETV 281
DB 485 -INSRQLRERVLFD---ELGLPAIGKTEKGRKSTSAVLEALREAHPIYEKILQYRELT 540
QY 282 EQ-----SPLXOKI--QTGRIDMOSVOT-----RLSDPAKGLDDRKRPYAGRLQDA 329
DB 541 KIKSYIDPLDLIRPTGRLEHTRFNQATATGRLSSDP--NLQNIPTPTPLGQRIRRA 598
QY 330 LMOPEKSVYLVSPY-----FVPTKSTDALAKLYODGIDVTVLTNS-----LQATDVA 377
DB 599 FLAEBEGLVALVYSOIELRVLALHLSGDENLIRVFOEGRDITHTASMMGVPREAVD-- 656
QY 378 AVHSGVYKRPDLKA-----GIKLYELOPNHAP----- 407
DB 657 -----FLMRAAKTINPGVLYGMSAHRLSQELAIPIYEAQAFIERYQSPKYR 705
QY 408 ----ATKDKGLGSSVTSIAHAKTFIVDGKRIFIGSFNLDPRSARLNTENGVIESP--K 460
DB 706 AMIETLTLEGGRRGYETLEF-----GRRRYVPLDEARVKSVEAER--MAFMNVQGT 757
QY 461 IAEQMERTLADTSP---EYAYRVTLDNRHRLQMHDPATKTYTNEPEALKMRIAIIIS 517

```

```

DB 758 AADLMKLAIVKLPRLIEENGARMLLOVHDELYLEAKERA-----EAVARLAKEVMEGVYP 813
QY 518 L-LPTE 522
DB 814 LAVPLE 819

RESULT 26
US-08-446-794A-2
Sequence 2, Application US/08446794A
Patent No. 5747327
GENERAL INFORMATION:
APPLICANT: DEKI, JUN
APPLICANT: MORIOKA, SHINJI
TITLE OF INVENTION: PHOSPHOLIPASE D GENE ORIGINATED FROM
TITLE OF INVENTION: PLANT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: P.O. BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,794A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0760-0203P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 812 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-446-794A-2

Query Match
Best Local Similarity 3.7%; Score 101; DB 1; Length 812;
Matches 108; Conservative 65; Mismatches 182; Indels 174; Gaps 26;

QY 61 IIOIRPTPTNGSIDLYLN-----DPHEAFARAALIESAHSLDLOYIWRNDI----- 111
DB 187 LYQDAIVP--DNFIPKPIPLADGKNYEPHRCMEDIFDAISMAQHLYITYGSSVTEITLVWD 245
QY 112 SGR-----LLENLYLAERGVRRLLL--DDNNTGRL--DGLLALD-----S 151
DB 246 SNRPKPGGVTLIGELKTKKASGVRLMLVWDDRTSVGLKRDGLMATDETEYVHGS 305
QY 152 HPIEIVRLNP-----FVLRKNRAGYLTDFPRLNRHMNKSFTADNR-----A 195
DB 306 DVNCVYLCPRNPDPDSGISIV-----QDLISITMFTH-----HOKIVVDHELPLNQSQQRRIV 356
QY 196 TILGSGNIDDEVFKVGEVDIVPADLIDLATGVSVGESHDFDRYWMASHAHNAATRIIRSG 254
DB 357 SFVGGGLDLDGGRDYQYHSLFTLD-----STHHD-----DFHDPNFASTIKKGG 402
QY 255 -----NICKGLQIA-----LGYNDET-----SRHALLRYRE-----TVQSPLYQKIQTG 293
DB 403 PREPMWIDHSRLRGLPIAMOVLYNFEQRMRKGGKDLILLQRLDLSITIIIPSPV--MFPED 460

```

```

Db 306 DVNQLCPRRNPDDSGSLV-----QDLSTMTFTH-----HOKIYVVDHELFPNOSSQQRRTV 356
QY 196 TLIGGRNIGDEYEFKVEDIVFADLDLIATGSSVGEVSHDFRYWASHAHNATRIIRSG- 254
Db 357 SFVGGDLDCGGRDYDTQYHSLFLRTLD-----STHHD-----DEHQPFATASIKKG 402
QY 255 -----NIGGLQA-----LGYNET-----SRHALLRRE-----TVESPLYQAIQNG 293
Db 403 PREPMWIDHRLLEPLIAMDVLYNEQRRKGGKDLQLLRDSDTITPPSPV--MPPED 460
QY 294 RIMQSVQTRBLISDDPAKGLDRDRKRPPVAG-----RLQD-----ALQPEKSVY 338
Db 461 RETMNVQLFRRSIDOGAAGFPDPPEEAKAGLVSGKQIIDRSIDQAYHAIIRAKNFYI 520
QY 339 LVSYTF-----VTKSGTDALAKL----- 357
Db 521 IENQVFGSSYAWKPEGIKPEDIGALHLIKELAKLVASKIEAGEERTYVVVPMMEGV 580
QY 358 -----VODGID-----VTVLTNSLQATDVAAVHSGVYK-----RKPLKAGIKLVE 399
Db 581 PESSVQAIIIDMQRRITMWMYTTDITELQAKGJEANPKDVLTFPCLGNRVKQAG-----E 636
QY 400 LQPNHVAVPATKDKGLGSS---VTSLHAKTFYIDGKRIFGTGSFNLDPRS 445
Db 637 YQPEQPEADTDYSRAQEARFRMIVYHTKMMIVDDEYIIIGSANINORS 685

RESULT 28
US-08-945-024-2
Sequence 2, Application US/08945024
Patent No. 5973226
GENERAL INFORMATION:
APPLICANT: Ueki, Jun
APPLICANT: MORIOKA, SHINJI
TITLE OF INVENTION: METHOD FOR CHANGING COMPOSITION OF
TITLE OF INVENTION: PHOSPHOLIPIDS PRODUCED BY ORGANISM AND RECOMBINANT VECTOR
TITLE OF INVENTION: THEREFOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/945,024
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 760-236
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ. ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 812 amino acids
type: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-945-024-2

Query Match 3.7%; Score 101; DB 2; Length 812;
Best Local Similarity 20.4%; Pred. No. 0.46;
Matches 108; Conservative 65; Mismatches 182; Indels 174; Gaps 26;

```

QY 61 LIQIRHTPTNGISDIYLLN---DPHEAPARAALIESAHSLSLDIYIIMRDI-----111  
 Db 187 LVQDAHVP--DNFIPKIPILADGKNPEPHRCWEDIFDAISNAQHILITYTMSVYTEITLVRD 245  
 QY 112 SGR-----LLENVYLAERGVRRLI--DNNNTRGL--DOLLALD-----S 151  
 Db 246 SNPKPGGVTLGELLKKAKEGVRMLVMDRTSVGLKRGMLATHEETENYFHGS 305  
 QY 152 HNIIEVRLFP-----FVLKRWALGYLTDPELRNRMHNSKFTADNR-----A 195  
 Db 306 DNVNVCYCPNRPDSSGISV---QDLSTMTFTH---HQIYVVDHELPNQSGQRRIY 356  
 QY 196 TILGNNIDGDEFFKVEDYVFADLDILATGSVYGEVSHDFDRKWSAHNATRIIRSG- 254  
 Db 357 SFVGGILCDGRYDQYHSLFETLD-----STHND---DEHOPMFATASIKKG 402  
 QY 255 -----NICKGLQA-----LGYNDET-----SRHALLRYRE---TVEOSPXYQKIQG 293  
 Db 403 PREPMHDIHSRLEGPILAMVLYNFEQRMKKGSKDILLQLRLSDTIIPSPV--WEPED 460  
 QY 294 RIDMOSVQTRLISDDPAKGLDRDRKPPIAG-----RLQD---ALKQPEKSVY 338  
 Db 461 RETMNVQLFRSIDGAAGFPDTPPEAKAGLVSGKQIIRSDIYIHAIRAKNFY 520  
 QY 339 LVSPLY-----VPTKSGTDALAKL-----357  
 Db 521 IENQYFLGSSYAMKPECIKPEDIGALHLPKELALKVSKSEAGEERTYVYVPMPEGV 580  
 QY 358 -----VQDQID-----VYLTNSLOATDYAAVHSGVYK---RKPLKAGIKLYE 399  
 Db 581 PESGSVALIDMOKRTMEMMYDITFALQAKGLEAMPKDYLTFFCLCNREVKQAG---E 636  
 QY 400 LQPNHAAVPAKDKGLTGS---VTSLHARTFYDGRKIFIGFENLDPRS 445  
 Db 637 YQPEQPEADTDYRSQAQEARREMYVHTKMIVYDEYIIIGSANINQRS 685

## RESULT 29

US-09-096-399-2  
 ; Sequence 2, Application US/09096399A  
 ; Patent No. 6130045  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wurst, Helmut  
 ; APPLICANT: Qui, Zhi-Hao  
 ; TITLE OF INVENTION: Thermostable Polymerase  
 ; FILE REFERENCE: CLON-007  
 ; CURRENT APPLICATION NUMBER: US/09/096,399A  
 ; CURRENT FILING DATE: 1998-06-11  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 553  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Recombinantly engineered mutant  
 US-09-096-399-2

Query Match 3.7%; Score 100; DB 4; Length 553;  
 Best Local Similarity 21.3%; Pred. No. 0.3;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVRV-----RLILDDNNTRGL--DOLLALDS---HPNIEVRLFNPF 163  
 Db 41 DLLALAAAGRGVHRAPEFYKALRDLKEARGLAKDLVLALEGLPFGDDPMI-----96  
 QY 164 VLKRWALGYLTD---FPLRNRBMHNSKFT--ADNRATI-----IGRNTGDE-----206  
 Db 97 -----LAVLDDPSNFTPEGVARRRGTEPAGERAAISERLFLANLWGLRGEERLLM 149  
 QY 207 -YKVGGE-----DTVFADLDILATGSVYGEVSHDFDRYAS-----HSAH 245  
 Db 150 LKREVERPLSAVLAHNEAGVRLDVAVYLRALSLEVAEELARLEAEVRLAGHPFNLSNR 209

QY 246 NATRIIRSGNICKGLQALGYNDETSRHA-----LTRYETVEQ-----283  
 Db 210 QLERVLFD---ELGLPAICKTEKTKGRSTSAANLEALREAHPIVEKILQYRELTFLKSNY 266  
 QY 284 -SPLYKI--QGRIDMOSVQ---RLISDDPAKGLDRDRKPPIAGRLQDALKQPK 335  
 Db 267 IDPLPDLIHRTRGLRHTRNQFATATGRSLSSDP--NLQNIIPVPTDLGQIRRAFAEESG 324  
 QY 336 SYLVESPY-----FVPTKSGTDALAKLVQDQIDYTVLNS-----LQATDYAAVHSGY 383  
 Db 325 WLVALDYQIELRVLAHLSGDENLIRVQEGSDITETASWMEGVPREAVD-----376  
 QY 384 VKYKPLRA-----GKLYELQPNHAP-----AT 409  
 Db 377 -----PLMRRAKTINFGVLYGMSAHLSQLAIPYEAQAFTERYFQSEPKVRAMIEKT 431  
 QY 410 KQKGLGSSVTSIHATFYDGRKIFIGSFNDRARLNTGCVYIESP---KINQOME 466  
 Db 432 LEEGRRGVYETLE-----GRRRYVPLEARVKSVERAER--KAFNMPVOGTADLAK 483  
 QY 467 RTLADTSP---EYAYRVTLDRHNRLOMHPATRKTYPNPEPAKIMRIAKILST-LPIE 522  
 Db 484 LAMVKLFPLREMGARMLQVHDELVLVLEAPKERA---EAVARLAKVEGYPPLAVPLE 539

## RESULT 30

US-08-021-623C-6  
 ; Sequence 6, Application US/08021623C  
 ; Patent No. 5436149  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barnes Ph.D., Wayne M.  
 ; TITLE OF INVENTION: Thermostable DNA polymerase with  
 ; TITLE OF INVENTION: Enhanced thermostability and enhanced length and  
 ; TITLE OF INVENTION: efficiency of primer extension  
 ; NUMBER OF SEQUENCES: 17  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Seuninger, Powers, Leavitt and Roedel  
 ; STREET: One Metropolitan Square, 16th Floor  
 ; CITY: St. Louis  
 ; STATE: Missouri  
 ; COUNTRY: USA  
 ; ZIP: 63102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk, 5.25", 360 kb.

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn 1.0, v1.25; EDIX; Wordperfect.

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/021,623C

FILING DATE: 19-FEB-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Blosser, G. Harley

REGISTRATION NUMBER: 33,650

REFERENCE/DOCKET NUMBER: WNB4900

TELECOMMUNICATION INFORMATION:

TELEPHONE: 314/231-5400

TELEFAX: 314/231-4342

TELEX: 6502697583 MCI

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 554 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-021-623C-6

Query Match 3.7%; Score 100; DB 1; Length 554;  
 Best Local Similarity 21.3%; Pred. No. 0.3;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVRV-----RLILDDNNTRGL--DOLLALDS---HPNIEVRLFNPF 163

```

Db      42 DDLALAAAGRGVHRAPEEPYKALRDLEKARGLAKDLVLALREGGLPPGDDPML----- 97
      164 VLKRMALGYLTD---FRLNRHMNKSFT--ADNRAT-----LGRNIGDE----- 206
      98 -----LAVLLDPNSNTTEGVAARRGGWTEBAGERALSERLFANLNGRLEGEERLLM 150
      207 -YFKVGE-----DYFADLDILATGSGVGEVSHDFDRYAS-----HSAH 245
      151 LYREVERPLSAVLAMHEATGVRDLVAYLRALSLEVAEELARLEAEVFRLAGHFNLNSRD 210
      246 NATRIIRSGNIGKLOALGYNDETSRA-----LRYRETEVO----- 283
      211 QLERVLFDD---ELGLPAIGTTEKTKRSTSAVLEALREAHPIVEKILOYRELTKLKSTY 267
      284 -SPLVOKI--OTGRIDMOSVOT-----RLISDDPAKGLDRDRKPPRAGRLDALKOPEK 335
      268 IDPLDPLHPRGRHLNTRNQTATATGRILSSSP--NLONIPVPTLGGIRIRAFIAEEG 325
      336 SVLVSPY-----FVPTKSGTDALAKLVODGIDVYLTNS-----LQATDVAVAHSGY 383
      326 WLVALDYSQIELRVLAHLSGDENLIRVQEGRDITETASMMFVGVPREAVD----- 377
      384 VYRKRPDLKA-----GKLYELOPNHAP-----AT 409
      378 -----PLMRAAKTINFGVLYGSAHRLSQELAIPEEAQAFIERYFQSPKVRAMIEKT 432
      410 KDKGLTGSVTSIHAKTFFIVDGKRIFFISFNLDPRSARLNTENGVIESP--KIAEOME 466
      433 LEEGRRGVETLTF-----GRRRYVPDLKRVKSVREAER--MAFNPVQGTADLKM 484
      467 RFLADTSP---EYAVRVTLDNRHRLQWMDPATRKTYPNPEAKLMKRIAKILSL-LPTIE 522
      485 LAMVKLFPRLLEKGARMLLQVHDELVLKPKERA---EAVARLAEVMEGVPLAVPLE 540

RESULT 31
US-09-096-399-4
; Sequence 4, Application US/09096399A
; Patent No. 6130045
; GENERAL INFORMATION:
; APPLICANT: Wurst, Helmut
; APPLICANT: Qui, Zhi-Hao
; TITLE OF INVENTION: Thermostable Polymerase
; FILE REFERENCE: CLON-007
; CURRENT APPLICATION NUMBER: US/09/096,399A
; CURRENT FILING DATE: 1998-06-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 559
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinantly engineered mutant
US-09-096-399-4

```

```

Query Match      3.7%; Score 100; DB 4; Length 559;
Best Local Similarity 21.3%; Pred. No. 0.31;
Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

```

```

QY      118 NUVYLAERGVN-----RLLDNNTRGL--DDLALADS---HPNIEVRLNPF 163
      47 DDLALAAAGRGVHRAPEEPYKALRDLEKARGLAKDLVLALREGGLPPGDDPML----- 102
      164 VLKRMALGYLTD---FRLNRHMNKSFT--ADNRAT-----LGRNIGDE----- 206
      103 -----LAVLLDPNSNTTEGVAARRGGWTEBAGERALSERLFANLNGRLEGEERLLM 155
      207 -YFKVGE-----DYFADLDILATGSGVGEVSHDFDRYAS-----HSAH 245
      156 LYREVERPLSAVLAMHEATGVRDLVAYLRALSLEVAEELARLEAEVFRLAGHFNLNSRD 215

```

```

QY      246 NATRIIRSGNIGKLOALGYNDETSRA-----LRYRETEVO----- 283
      216 QLERVLFDD---ELGLPAIGTTEKTKRSTSAVLEALREAHPIVEKILOYRELTKLKSTY 272
      284 -SPLVOKI--OTGRIDMOSVOT-----RLISDDPAKGLDRDRKPPRAGRLDALKOPEK 335
      273 IDPLDPLHPRGRHLNTRNQTATATGRILSSSP--NLONIPVPTLGGIRIRAFIAEEG 330
      336 SVLVSPY-----FVPTKSGTDALAKLVODGIDVYLTNS-----LQATDVAVAHSGY 383
      331 WLVALDYSQIELRVLAHLSGDENLIRVQEGRDITETASMMFVGVPREAVD----- 382
      384 VYRKRPDLKA-----GKLYELOPNHAP-----AT 409
      383 -----PLMRAAKTINFGVLYGSAHRLSQELAIPEEAQAFIERYFQSPKVRAMIEKT 437
      410 KDKGLTGSVTSIHAKTFFIVDGKRIFFISFNLDPRSARLNTENGVIESP--KIAEOME 466
      438 LEEGRRGVETLTF-----GRRRYVPDLKRVKSVREAER--MAFNPVQGTADLKM 489
      467 RFLADTSP---EYAVRVTLDNRHRLQWMDPATRKTYPNPEAKLMKRIAKILSL-LPTIE 522
      490 LAMVKLFPRLLEKGARMLLQVHDELVLKPKERA---EAVARLAEVMEGVPLAVPLE 545

```

```

RESULT 32
US-07-977-434-2
; Sequence 2, Application US/07977434
; Patent No. 546591
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: wordperfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977,434
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 746,121

```

FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US90/07641  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 585,471  
FILING DATE: 20-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 455,611  
FILING DATE: 22-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 609,157  
FILING DATE: 02-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 557,517  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Luann Caser  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 814-2972  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-977-434-2

Query Match 3.7%; Score 100; DB 1; Length 832;  
Best Local Similarity 21.3%; Pred. No. 0.61;  
Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVR-----RLLDNNRGL---DDLALDS---HPNIEVLFNPF 163  
DB 320 DDLALAAAGGVHRAPEYKALRDKEARGLAKDLVSLALREGLGPPGDDPHL---- 375  
QY 164 VLRKRALGYLTD---FPLNRMRHNSFT---ADNRATI-----LGRNIGDE---- 206  
DB 376 -----LAVLLDPNSTPPEGVARRYGGEWTEAGEBRALSRLFANLWGRLEGEBRLM 428  
QY 207 -YFKVGE-----DTVFADLLATGSGVYGEVSHDFRYWAS-----HSAH 245  
DB 429 LYREVERPLSAVLAMHATGVRLDVAYLRLSLVEAEIARLEAEVFRLAGHPFNLSRD 488  
QY 246 NATRIIRSGNICKGLQALGYNDTSRHA-----LTRYRETVQ----- 283  
DB 489 QERVLFD---ELGLPAICKTEKTKRSTSAVALEALREAHPIVEKILQYRELTKLSTY 545  
QY 284 -SPLYOKI--QTRIDWOSVQT-----RLISDPAKGLDRRKPPIAGRLDALKOPEK 335  
DB 546 IDPLDLHLHPTGRILHTRNQATATATGRLSSDP--NLQNTIVFRPPLGGRIRRAIAEG 603  
QY 336 SVYLVSFY-----FVPTKSGTDALAKLVQDIDVTVLNS-----LQATVAAVHSGY 383  
DB 604 WLVLALDYSQILRVLAHLSGDENLIRVFOESGRDITHTETASMMFGVPREAVD----- 655  
QY 384 VKYRKPPLKA-----GIKLYELOPNHVP-----AT 409  
DB 656 -----PLMRRAKTIINFGVLVGMSAHRLSQELAIPEEQAQIERFYFOSFPKVRAMIEKT 710  
QY 410 KDKGLTSSSVTSLHAKTIVDGKRIFFIGSFNDPRSARLNTMGVYIESP---KIAEQME 466  
DB 711 LEEGRRRGIVETLF-----GRRRYVPDLERAVKSVREAAER--MAFNMPVGGTAADLMK 762  
QY 467 RFLADTSP---EYARVYLDLRHNRQWMDPARRKYTPNPEPKLKKRIAKLISL-LPIE 522  
DB 763 LAMVKLFPRLHEMGARMLQVHDELVAEKERA---EAVARLAKVEGVPPLAVPLE 818

RESULT 33  
US-08-156-020-2

Sequence 2, Application US/08156020  
Patent No. 5474920  
GENERAL INFORMATION:  
APPLICANT: Moses M.D., Robb E.  
TITLE OF INVENTION: Modified Thermo-Resistant DNA  
TITLE OF INVENTION: Polymerases  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Allegretti & Witcoff  
STREET: 10 South Wacker Drive  
CITY: Chicago  
STATE: IL  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/156,020  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Greenfield Ph.D., Michael S.  
REGISTRATION NUMBER: 37,142  
REFERENCE/DOCKET NUMBER: 93,413  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312)715-1000  
TELEFAX: (312)715-1234  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 832 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-156-020-2

Query Match 3.7%; Score 100; DB 1; Length 832;  
Best Local Similarity 21.3%; Pred. No. 0.61;  
Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVR-----RLLDNNRGL---DDLALDS---HPNIEVLFNPF 163  
DB 320 DDLALAAAGGVHRAPEYKALRDKEARGLAKDLVSLALREGLGPPGDDPHL---- 375  
QY 164 VLRKRALGYLTD---FPLNRMRHNSFT---ADNRATI-----LGRNIGDE---- 206  
DB 376 -----LAVLLDPNSTPPEGVARRYGGEWTEAGEBRALSRLFANLWGRLEGEBRLM 428  
QY 207 -YFKVGE-----DTVFADLLATGSGVYGEVSHDFRYWAS-----HSAH 245  
DB 429 LYREVERPLSAVLAMHATGVRLDVAYLRLSLVEAEIARLEAEVFRLAGHPFNLSRD 488  
QY 246 NATRIIRSGNICKGLQALGYNDTSRHA-----LTRYRETVQ----- 283  
DB 489 QERVLFD---ELGLPAICKTEKTKRSTSAVALEALREAHPIVEKILQYRELTKLSTY 545  
QY 284 -SPLYOKI--QTRIDWOSVQT-----RLISDPAKGLDRRKPPIAGRLDALKOPEK 335  
DB 546 IDPLDLHLHPTGRILHTRNQATATATGRLSSDP--NLQNTIVFRPPLGGRIRRAIAEG 603  
QY 336 SVYLVSFY-----FVPTKSGTDALAKLVQDIDVTVLNS-----LQATVAAVHSGY 383  
DB 604 WLVLALDYSQILRVLAHLSGDENLIRVFOESGRDITHTETASMMFGVPREAVD----- 655  
QY 384 VKYRKPPLKA-----GIKLYELOPNHVP-----AT 409  
DB 656 -----PLMRRAKTIINFGVLVGMSAHRLSQELAIPEEQAQIERFYFOSFPKVRAMIEKT 710  
QY 410 KDKGLTSSSVTSLHAKTIVDGKRIFFIGSFNDPRSARLNTMGVYIESP---KIAEQME 466  
DB 711 LEEGRRRGIVETLF-----GRRRYVPDLERAVKSVREAAER--MAFNMPVGGTAADLMK 762

OY 467 RFLADTSP---EYAYRTLDNRHNLQWHDPAETKTYNPEEAKLWKRIAKIISL-LPIE 522  
DB 763 LAMVKLFPRLEMGARMLQVHDELVLAEPKERA-----EAVARLAEVMEGVPLAVPLE 818

## RESULT 34

US-08-156-020-4  
Sequence 4, Application US/08156020

Patent No. 5474920  
GENERAL INFORMATION:

APPLICANT: Moses M.D., Robb E.

TITLE OF INVENTION: Modified Thermo-Resistant DNA

NUMBER OF INVENTION: Polymerases

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti & Witcoff

STREET: 10 South Wacker Drive

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/156,020

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.

REGISTRATION NUMBER: 37,142

REFERENCE/DOCKET NUMBER: 93,413

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312)715-1000

TELEFAX: (312)715-1234

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-156-020-4

Query Match 3.7%; Score 100; DB 1; Length 832;

Best Local Similarity 21.3%; Pred. No. 0.61;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

OY 118 NLVYLAERGVY-----RLLDNNTRGL---DDLALDS---HPNIEVRLNPF 163  
DB 320 DLALAAAGRGVRAPEPKALRDLKEARGLAKDLISVALRGLGPGDDPML----- 375  
OY 164 VLKRWALGYLTD---FPLNRMHKSF--ADNRATI-----LGGNIGDE----- 206  
DB 376 -----LAVLLDPSNTTPGVARRYGGEWTEEGERAALSERLFANLWGRLEGEERLLW 428  
OY 207 -YKVG-----DTVFADLDIATGVSVEGSHDFRYMAS-----HSAH 245  
DB 429 LYREVERPLSAVLAMHATGVRLDVALRALSLAEVRIARLEAEVRLAGHPFNLSRD 488  
OY 246 NARRIRSGNIGKGLQALGNDERSHA-----LRRREYEO----- 283  
DB 489 QLERVLPD---ELGLPAIGTKTEGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTY 545  
OY 284 -SPYKRI--QTGRIDQSVQV-----RLISDPKAGLDRDRKPPAGRLQDALOKPEK 335  
DB 546 IDPLPDLIHRTGRHLHRTMOTATATGRLLSSDP--NLQNIPTPTPLGQRIKRAFIABEG 603  
OY 336 SYLVASFY-----EVPFKSGDALAKLVODGIDVTVLNS-----LQATVAVAHSGY 383  
DB 604 MLIVALDYSOIELRLVLAHLTSGDENLIRVFOGBRDIHTEFTASWMEGVPREAVD----- 655

OY 384 VKYRKPLKA-----GIKIVELQPNHAP-----AT 409

DB 656 -----PLMKRAATINFGVLYGSAHLSOELAIPIEEAOFTERTQSPKRAMTEKT 710

OY 410 KDKGLTGSVYSLHAKTFIVDGKRIPIGSEFNLDPRSARLNTENGVIESP---KIAOME 466

DB 711 LEEGRRGYVETLF-----GRRRYVPDLKARVKSVEAAER-MAFNMPVQGTADLMK 762

OY 467 RFLADTSP---EYAYRTLDNRHNLQWHDPAETKTYNPEEAKLWKRIAKIISL-LPIE 522  
DB 763 LAMVKLFPRLEMGARMLQVHDELVLAEPKERA-----EAVARLAEVMEGVPLAVPLE 818

## RESULT 35

US-08-156-020-6

Sequence 6, Application US/08156020

Patent No. 5474920

GENERAL INFORMATION:

APPLICANT: Moses M.D., Robb E.

TITLE OF INVENTION: Modified Thermo-Resistant DNA

NUMBER OF INVENTION: Polymerases

CORRESPONDENCE ADDRESS:

ADDRESSEE: Allegretti & Witcoff

STREET: 10 South Wacker Drive

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/156,020

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.

REGISTRATION NUMBER: 37,142

REFERENCE/DOCKET NUMBER: 93,413

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312)715-1000

TELEFAX: (312)715-1234

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-156-020-6

Query Match 3.7%; Score 100; DB 1; Length 832;

Best Local Similarity 21.3%; Pred. No. 0.61;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

OY 118 NLVYLAERGVY-----RLLDNNTRGL---DDLALDS---HPNIEVRLNPF 163  
DB 320 DLALAAAGRGVRAPEPKALRDLKEARGLAKDLISVALRGLGPGDDPML----- 375  
OY 164 VLKRWALGYLTD---FPLNRMHKSF--ADNRATI-----LGGNIGDE----- 206  
DB 376 -----LAVLLDPSNTTPGVARRYGGEWTEEGERAALSERLFANLWGRLEGEERLLW 428  
OY 207 -YKVG-----DTVFADLDIATGVSVEGSHDFRYMAS-----HSAH 245  
DB 429 LYREVERPLSAVLAMHATGVRLDVALRALSLAEVRIARLEAEVRLAGHPFNLSRD 488  
OY 246 NARRIRSGNIGKGLQALGNDERSHA-----LRRREYEO----- 283  
DB 489 QLERVLPD---ELGLPAIGTKTEGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTY 545

QY 284 -SPLYKTI--QTGRIDMOSVQT-----RLISDDPAKGLDRRKPPIAGRLDALKOPK 335  
DB 546 IDPLPDLIHRGRRLTRRQTATGRSLSSDP--NLONIPVTRTGLGRIRAFIAEAG 603  
QY 336 SYLVSPY-----FVPTSGTDALAKLVQDGDVTLVNS-----LQATDVAVAHSGY 383  
DB 604 WLLVALDYQIQLRVLAHLSGDENLIRFQEGHDITHTASWFGVPREAVD----- 655  
QY 384 VYKRKPLKA-----GIKLYELQPNHAVP-----AT 409  
DB 656 -----PLMRRAKTIINGVLYGSAHRLSGELAIPEEAQFIERYFQSPKVRAMIEKT 710  
QY 410 KKGGLTSSVTSIAKTFIVDGKRIFGSPNLDPRSARLNTENGVIESP---KIAEOME 466  
DB 711 LEEGRRGVETLF-----GRRRYVPDLKARVKSVRKAAER--MAFNMPVQGTADLAK 762  
QY 467 RLADTSP---EYAYRVTLDRHNRLOMHPATRKTYPNPEAKLWKRIAKILSL-LPIE 522  
DB 763 LAMVKLFPRLEEGAMRLQVHDELVLKPKERA---EAVARLAKVEGVIPLAVPLE 818

## RESULT 36

US-08-156-020-8  
Sequence 8, Application US/08156020

Patent No. 5474920

GENERAL INFORMATION:

APPLICANT: Moses M.D., Robb E.

TITLE OF INVENTION: Modified Thermo-Resistant DNA

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 10 South Wacker Drive

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/156,020

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.

REGISTRATION NUMBER: 37,142

REFERENCE/DOCKET NUMBER: 93,413

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312)715-1000

TELEFAX: (312)715-1234

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-156-020-8

Query Match 3.7%; Score 100; DB 1; Length 832;

Best Local Similarity 21.3%; Pred. No. 0.61; Indels 176; Gaps 27;

Matches 115; Conservative 66; Mismatches 183;

QY 118 NLVYLAERGVR-----RLLLDNNTRGL---DDLALDS---HPNIEVRLNPF 163  
DB 320 DDLAALAAARGVHRAPERYKALRDLKEARGLLADLSVLALREGIGLPPGDPMI----- 375  
QY 164 VLRKRALGYLTD-----FPLNRMRHNSFT--ADNRATI-----LGRNIGDE--- 206  
DB 376 -----LAVLLDPSNTPEGVARRYGEMTEAGEERALISRLPANLWGRLEGEERLIM 428

QY 207 -YFKVE-----DTVFADLLIATGVSVEVSHDEDRYAS-----HSAH 245  
DB 429 LYREVERPLSAVLAHMEAGVRLDVAYLRLSLVVAEELARLEAEVRLAGHPFNLSND 488  
QY 246 NATRIIRSGNIGGLQALGYNDETSRHA-----LKYREVEO----- 283  
DB 489 QLETVLFD---ELGLPAIGKTEKTRKRSAAVLEALREAHPIVERKILQYRELTKLKSTY 545  
QY 284 -SPLYKTI--QTGRIDMOSVQT-----RLISDDPAKGLDRRKPPIAGRLDALKOPK 335  
DB 546 IDPLPDLIHRGRRLTRRQTATGRSLSSDP--NLONIPVTRTGLGRIRAFIAEAG 603  
QY 336 SYLVSPY-----FVPTSGTDALAKLVQDGDVTLVNS-----LQATDVAVAHSGY 383  
DB 604 WLLVALDYQIQLRVLAHLSGDENLIRFQEGHDITHTASWFGVPREAVD----- 655  
QY 384 VYKRKPLKA-----GIKLYELQPNHAVP-----AT 409  
DB 656 -----PLMRRAKTIINGVLYGSAHRLSGELAIPEEAQFIERYFQSPKVRAMIEKT 710  
QY 410 KKGGLTSSVTSIAKTFIVDGKRIFGSPNLDPRSARLNTENGVIESP---KIAEOME 466  
DB 711 LEEGRRGVETLF-----GRRRYVPDLKARVKSVRKAAER--MAFNMPVQGTADLAK 762  
QY 467 RLADTSP---EYAYRVTLDRHNRLOMHPATRKTYPNPEAKLWKRIAKILSL-LPIE 522  
DB 763 LAMVKLFPRLEEGAMRLQVHDELVLKPKERA---EAVARLAKVEGVIPLAVPLE 818

## RESULT 37

US-08-156-020-10

Sequence 10, Application US/08156020

Patent No. 5474920

GENERAL INFORMATION:

APPLICANT: Moses M.D., Robb E.

TITLE OF INVENTION: Modified Thermo-Resistant DNA

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

STREET: 10 South Wacker Drive

CITY: Chicago

STATE: IL

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Apple Macintosh

OPERATING SYSTEM: Macintosh

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/156,020

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Greenfield Ph.D., Michael S.

REGISTRATION NUMBER: 37,142

REFERENCE/DOCKET NUMBER: 93,413

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312)715-1000

TELEFAX: (312)715-1234

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 832 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-156-020-10

Query Match 3.7%; Score 100; DB 1; Length 832;

Best Local Similarity 21.3%; Pred. No. 0.61;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;





OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/254,359A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/073,384  
 FILING DATE: 06-JUN-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/986,330  
 FILING DATE: 07-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: CARROLL, PETER G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: FORS-01000  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ. ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 832 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-254-359A-4

Query Match 3.7%; Score 100; DB 1; Length 832;  
 Best Local Similarity 21.3%; Pred. No. 0.61;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVR-----RLTDDNNTGRL--DGLLALDS-----HPNIEVRLFNPF 163  
 DB 320 DILALAAAGGRVHRAPEPKALRDLEKARGLLAKLSVALKEGIGLPPGDDPML----- 375  
 QY 164 VLRKRALGYLD---FPRLNRMHNSPT--ADNRAT-----LGRNIGDE----- 206  
 DB 376 -----LAVLLDPSNTTPEGVARYGGEWTEAGERALSERLFANLWGRLEGERLLM 428  
 QY 207 -YFKVGE-----DIYFADIDILATGVSVEGSHDFDRYAS-----HSAH 245  
 DB 429 LYREVERPLSAVLAHMEATGVRDLVAYLRALSLAEVAELARLEAEVFRLAGHPFNLSRD 488  
 QY 246 NATRIIRSGNIGKQALGYNDETSRHA-----LTRYRETVED----- 283  
 DB 489 QLERVLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIVEKILQIRELTCLKSTY 545  
 QY 284 -SPLYOKI--QGRIDMOSVQT-----RLISDPAKGLDRDRRKPPIAGRLQDALQOPK 335  
 DB 546 IDPLPLIHPRGRLHTRNOTATATGRSSSDP--NLQNIYPRPLGGRIRAFIAEGS 603  
 QY 336 SYLVSPY-----EYPTSGTDALAKLVQDGDIVTVLNS-----LOATDVAAVHSGY 383  
 DB 604 WLVLVADYSQIELRVLAHLSGDENLIRVQEGRDITHTASWMEGVPRRAVD----- 655  
 QY 384 VYRKPLKA-----GIKLYELQPNHNAV-----AT 409  
 DB 656 -----PLMRRAKTIINEGVLYGMSAHLRSOELAIPIYEEQALPIERYFOSEPVRANIEKT 710  
 QY 410 KDKGLTGSSVSTLHAKFTIVDKRIFIGSFNLDPDSARLNTENGVIYESP--KIAEQME 466  
 DB 711 LREGRRRGVETLLE-----GRRRYVPLLEARVKSVRRAER--MAFMNPFVQGTADLAK 762  
 QY 467 RLADTSP---EYAYRVTLDRHNRLQWMDPATRKTYPNPEPEAKLWKRIAKLITL-LPIE 522  
 DB 763 IAMVKLFPLREMGAMLLQVHDELVEAPKERA---EAVARLAKEVMEGVPLAVPE 818

RESULT 40  
 US-08-483-043-4  
 ; Sequence 4, Application US/08483043  
 ; Patent No. 5691142

GENERAL INFORMATION:  
 APPLICANT: Dahlberg, James E.  
 APPLICANT: Lyamichev, Victor I.  
 APPLICANT: Brow, Mary Ann D.  
 TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA  
 TITLE OF INVENTION: POLYMERASE  
 NUMBER OF SEQUENCES: 29  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL  
 STREET: 220 Montgomery Street, Suite 2200  
 City: San Francisco  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/483,043  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/073,384  
 FILING DATE: 04-JUN-1993  
 APPLICATION NUMBER: US 07/986,330  
 FILING DATE: 07-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Carroll, Peter G.  
 REGISTRATION NUMBER: 32,837  
 REFERENCE/DOCKET NUMBER: FORS-00613  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415/705-8410  
 TELEFAX: 415/397-8338  
 INFORMATION FOR SEQ. ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 832 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-483-043-4

Query Match 3.7%; Score 100; DB 1; Length 832;

Best Local Similarity 21.3%; Pred. No. 0.61;

Matches 115; Conservative 66; Mismatches 183; Indels 176; Gaps 27;

QY 118 NLVYLAERGVR-----RLTDDNNTGRL--DGLLALDS-----HPNIEVRLFNPF 163  
 DB 320 DILALAAAGGRVHRAPEPKALRDLEKARGLLAKLSVALKEGIGLPPGDDPML----- 375  
 QY 164 VLRKRALGYLD---FPRLNRMHNSPT--ADNRAT-----LGRNIGDE----- 206  
 DB 376 -----LAVLLDPSNTTPEGVARYGGEWTEAGERALSERLFANLWGRLEGERLLM 428  
 QY 207 -YFKVGE-----DIYFADIDILATGVSVEGSHDFDRYAS-----HSAH 245  
 DB 429 LYREVERPLSAVLAHMEATGVRDLVAYLRALSLAEVAELARLEAEVFRLAGHPFNLSRD 488  
 QY 246 NATRIIRSGNIGKQALGYNDETSRHA-----LTRYRETVED----- 283  
 DB 489 QLERVLFD---ELGLPAIGKTEKTKRSTSAVLEALREAHPIVEKILQIRELTCLKSTY 545  
 QY 284 -SPLYOKI--QGRIDMOSVQT-----RLISDPAKGLDRDRRKPPIAGRLQDALQOPK 335  
 DB 546 IDPLPLIHPRGRLHTRNOTATATGRSSSDP--NLQNIYPRPLGGRIRAFIAEGS 603  
 QY 336 SYLVSPY-----EYPTSGTDALAKLVQDGDIVTVLNS-----LOATDVAAVHSGY 383  
 DB 604 WLVLVADYSQIELRVLAHLSGDENLIRVQEGRDITHTASWMEGVPRRAVD----- 655  
 QY 384 VYRKPLKA-----GIKLYELQPNHNAV-----AT 409

Db 656 ----PIMRRAKTINGVLYGMSAHRLSOELAIPEEAQAFIERYFQSEFPKVRANIEKT 710  
QY 410 KOKGLTGSSVTSIHAFTIVDGKRIFIGSFNLDPRSARLNTENGVIESP---KIAEQME 466  
Db 711 LEEGRRRGVETLE-----GRRRYVPDLEARVKSVEREAER-MAFNMPVOGTADLMK 762  
QY 467 RFLADTSP--EYAVRVTLDRHNRLOMHPATRKTYPNEPEAKLMKRIAKIISL-LPIE 522  
Db 763 LAMVKLFPRLEENGARMLLOVHDELVLLEAPKERA---EAVARLAKKEVMEGVYPLAVPLE 818

Search completed: May 12, 2003, 10:01:06  
Job time : 42 secs



Db 117 DNDLDTNFLDLDALFEGVASKAKESFERYWRFHRSIPVSLIRT----- 160

OY 269 TSHALLKREVEOSPLYOKIOTGRIDMOSVOTRL-----ISD 307  
DB 161 ---HRRK--NNAKEIATLHEKIPISAEKNOKEFKYNDIFDRFOKYQYPIYYNAIFLAD 216  
OY 308 DPAKGLDRRRKPIAGRLQDALKOPEKSVLYSPYFVPTKSGTALAKLVDGIDVTL 367  
DB 217 SPKK-IDIPPLY-PIKIAFEKALKKAKOSVFIASTFFIGKMMKIFKQIKSGIELNLT 274  
OY 368 TNSLATDVAVAHSGVYKRPPLKAGIKLYELOPHAVAPKDKGLTGSVTSIAKTF 427  
DB 275 TNSLSTDAIVYVYGAHERNQLVIRMGANVYEIRNDFEQRQKGR---FSTKHSIAKTI 331  
OY 428 IYDGKRIFGSNLDPKRSARLTENGVIYESPKIAQOMERTLADTSPEYAVYTLDRNR 487  
DB 332 VEDMLTLLGSEFNIDPSRYINTESAVLEFDNPSFAKRVRLSLKDHA-QOSWMLVYVRH-R 389  
OY 488 LQW-----HDPATKTYNEPEAKLMKRIAKIISLPLESI 524  
DB 390 VYWEAVEGILHE---KT---SPDTSEFLLIKEMSKVLEPEREL 428

## RESULT 2

US-09-828-523A-22  
; Sequence 22, Application US/09828523A  
; Patent No. US20020168697A1  
; GENERAL INFORMATION:  
; APPLICANT: The Pharmacia & Upjohn Company  
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS  
; FILE REFERENCE: 268.62120101  
; CURRENT APPLICATION NUMBER: US/09/828,523A  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/266,327  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 99  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 493  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-828-523A-22

Query Match 10.1%; Score 275; DB 9; Length 493;  
Best Local Similarity 21.5%; Pred. No. 4,4e-16;  
Matches 100; Conservative 81; Mismatches 149; Indels 136; Gaps 16;

OY 74 SDIY--LNDPHEAFARAALIESAHSIDLOYYIWRNDISGRLLFNLYLAERGVR 131  
DB 139 NDLYDQVLKD-----IKNAKEYIHLEYTFALDGLGRILHAEELKQGLEVK 187  
OY 132 LLLDNNTRGDLDDLALDSHNIIEVRLFPVLKRWALG-----YLTDP---RLN 181  
DB 188 ILYDVGSK-----NVKMANPDHF-----KSLGEVEAFPSKLPILNFRMN 229  
OY 182 RRMHNSFTADNRATILGRNIGDEYFKVGEDTVEFADIIATGSGVGEVSHDFPRYAS 241  
DB 230 NNNHKKIIVIDQOLGYGVGFNGIDELGIGKIGYWRDTHLRIGQGAVALQALRFLIDMNS 289  
OY 242 HSAHNATRII-----RSGNIGKGLQALGYNDETSRHALLRYRETVEOSPLYOKIOTGR 294  
DB 230 -QAHNPOFEYDVKKYPPKKNPGLG-----NSPIQIAASGPA 323  
OY 295 IDMOVSQV---FRLISDDPAKGLDRRRKPIAGRLQDALKOPEKSVLYSPYFVPTKSGT 351  
DB 324 SDMHQIEGYTKMT-----MSAKKSVYLOSFPFIDNSYI 358  
OY 352 DALAKLVODGIDVTLTNSLATDVAVAHSGVYKRPPLKAGIKLYELOPHAVAPKDK 411  
DB 359 NAIKIAASGVYVHLMIC--KPDHPLVYWFATFSNASDLSSGVKTY-----TYE 406  
OY 412 KGLTGSVTSIAKTFIYDGKRIFGSNLDPKRSARLTENGVIYESPKIAQOMERTLAD 471  
DB 407 NGF-----IHSKMCLIDDEIVSGTANMDRSEFELNEVNAFYVDENLAKDL----- 453

OY 472 TSPKAYRVYTLDRHNRQLQWHDPAKTRKTYNEPEAKLMKRIAKIIS 517  
DB 454 ---RVAVEHDITKSKOL-----TKESYANRPLSVKFKRESIAKTIVS 490

## RESULT 3

US-09-828-523A-90  
; Sequence 90, Application US/09828523A  
; Patent No. US20020168697A1  
; GENERAL INFORMATION:  
; APPLICANT: The Pharmacia & Upjohn Company  
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS  
; FILE REFERENCE: 268.62120101  
; CURRENT APPLICATION NUMBER: US/09/828,523A  
; CURRENT FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: 60/266,327  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 99  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 90  
; LENGTH: 502  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: amino acid sequence encoded by S. aureus coding region cloned  
US-09-828-523A-90

Query Match 10.1%; Score 275; DB 9; Length 502;  
Best Local Similarity 21.5%; Pred. No. 4,6e-16;  
Matches 100; Conservative 81; Mismatches 149; Indels 136; Gaps 16;

OY 74 SDIY--LNDPHEAFARAALIESAHSIDLOYYIWRNDISGRLLFNLYLAERGVR 131  
DB 140 NDLYDQVLKD-----IKNAKEYIHLEYTFALDGLGRILHAEELKQGLEVK 188  
OY 132 LLLDNNTRGDLDDLALDSHNIIEVRLFPVLKRWALG-----YLTDP---RLN 181  
DB 189 ILYDVGSK-----NVKMANPDHF-----KSLGEVEAFPSKLPILNFRMN 230  
OY 182 RRMHNSFTADNRATILGRNIGDEYFKVGEDTVEFADIIATGSGVGEVSHDFPRYAS 241  
DB 231 NNNHKKIIVIDQOLGYGVGFNGIDELGIGKIGYWRDTHLRIGQGAVALQALRFLIDMNS 290  
OY 242 HSAHNATRII-----RSGNIGKGLQALGYNDETSRHALLRYRETVEOSPLYOKIOTGR 294  
DB 291 -QAHNPOFEYDVKKYPPKKNPGLG-----NSPIQIAASGPA 324  
OY 295 IDMOVSQV---FRLISDDPAKGLDRRRKPIAGRLQDALKOPEKSVLYSPYFVPTKSGT 351  
DB 325 SDMHQIEGYTKMT-----MSAKKSVYLOSFPFIDNSYI 359  
OY 352 DALAKLVODGIDVTLTNSLATDVAVAHSGVYKRPPLKAGIKLYELOPHAVAPKDK 411  
DB 360 NAIKIAASGVYVHLMIC--KPDHPLVYWFATFSNASDLSSGVKTY-----TYE 407  
OY 412 KGLTGSVTSIAKTFIYDGKRIFGSNLDPKRSARLTENGVIYESPKIAQOMERTLAD 471  
DB 408 NGF-----IHSKMCLIDDEIVSGTANMDRSEFELNEVNAFYVDENLAKDL----- 454  
OY 472 TSPKAYRVYTLDRHNRQLQWHDPAKTRKTYNEPEAKLMKRIAKIIS 517  
DB 455 ---RVAVEHDITKSKOL-----TKESYANRPLSVKFKRESIAKTIVS 491

## RESULT 4

US-09-738-626-6514  
; Sequence 6514, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO



27:

•

```

QY 105 YWRNDISRLFNLYLAERGVRLLDNNRGLDLDLAL--DSHPNIEVRLP- 160
DB 66 ---QNNPNRMIICDRI-ROAORIGIFMRIY-----TISSDIIOSLIQTSHHVVEVKYHC 116
QY 161 ---NPEVLKRWALGVLTDEPRLRMRHNSKSPADNRATIIIGRNIGDEYFKGEPTVA 217
DB 117 GESLPACONSRKRYVLRITNGRTL--QHKTKMLADQTVVGTGSAN-----YT 160
QY 218 DLDIATGVSVEVSHDFDRYMASSHANATRIIRSGNIGKGLQALGYNDSTRHALLRY 277
DB 161 DLSL-----NHD-----ANVTACI----- 174
QY 278 RETVEGSPYKIQIGRIMOSVQTRILISDDPAKGLDRRKRPIAGR-LQDALKOPKS 336
DB 175 ---ESSELHDAVSESRPOLVHGPQLNVIPIQ-----RLIPNASKMILNAINQATDS 225
QY 337 VYL-----VSPYFVPTKSGTDALAKLVODIDVTVLNLSQATDVAHVSGYKRPPL 391
DB 226 IFVLMYIFLSPFEL-----ALAQAMRGVARKYIIDHNSKODIC-----KLLS 269
QY 392 KAGIKI--YELDPNNAVPATKDKGLTGSVTSIHAFTYVDGKRIFIGSFNLDPRSARLN 449
DB 270 KGIOLPIYE-----RKTEGV-----LHFKICCIDMKKTLIFGSANNSGAGMIRN 313
QY 450 TEMGVYESPKIAEOMER-----TLADTSPEYAYRVTLDNRHNLQWHDPAETKT 498
DB 314 FE-DLFIPLRPIETQLOAFMDVMSLETNNSY-----LSPEVLTAPTPSSRPT 361

```

# RESULT 7

US-10-033-297-69  
 Sequence 69, Application US/10033297  
 Publication No. US20020187486A1

## GENERAL INFORMATION:

APPLICANT: Hall, Jeff G.  
 Lyamichev, Victor I.  
 Mast, Andrea L.

TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
 Sequential Invasive Cleavages

NUMBER OF SEQUENCES: 163  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: Medien & Carroll, LLP  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States Of America  
 ZIP: 94104

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/10/033,297  
 FILING DATE: 12-NOV-2001  
 CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/350,597  
 FILING DATE: 09-JUL-1999  
 APPLICATION NUMBER: US/08/823,516  
 FILING DATE: 24-MAR-1997  
 APPLICATION NUMBER: PCT/US97/01072  
 FILING DATE: 21-JAN-1997  
 APPLICATION NUMBER: US 08/759,038  
 FILING DATE: 02-DEC-1996  
 APPLICATION NUMBER: US 08/758,314  
 FILING DATE: 02-DEC-1996  
 APPLICATION NUMBER: US 08/756,386  
 FILING DATE: 29-NOV-1996  
 APPLICATION NUMBER: US 08/662,853  
 FILING DATE: 12-JUL-1996  
 APPLICATION NUMBER: US 08/599,491

```

FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 833 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 69:
US-10-033-297-69

```

Query Match: 3.9%; Score 106; DB 9; Length 833;  
 Best Local Similarity: 21.8%; Pred. No. 1.2; Indels 188; Gaps 30;  
 Matches 119; Conservative 65; Mismatches 174;

```

QY 118 NLVYLAERGVAV-----RLLDNNRGL-----DLDLALDS-----HPNIEVRLNPF 163
DB 321 DLLALAAAGRVHRAPEYKALRDLEKARGLADSLALREGLGPPGDDPYL----- 376
QY 164 VRKMRALCYLTD-----PRLNRMRHNSFT--ADNRATI-----LGRNIGDE----- 206
DB 377 -----LAVLLDPSNTTPEGVARRYGGEWTEAGERALSERLPANLNGRLEGGERLIM 429
QY 207 -YFKYGE--DIVEADLDILANG-----SYNGEVS-----HFDKRYM 239
DB 430 LYREVEPLSAVLAHME--ATGVRLDVAVALRLSLEVAGETARLEAVEVRLAGHPFN--- 484
QY 240 ASHSANMATRIIRSGNIGKGLQALGYNDSTSRHA-----LARKRETV 281
DB 485 -LNSDQLELVFD--ELGIPAIKTEKTKGRNSAAVLEALRAHPIVEKILQYREL 540
QY 282 EQ-----SPLVQKI--QGRIMOSVQTR--RLISDDPAKGLDRRKRPIAGRLQDA 329
DB 541 KUKSTYIDPLDPLHPRGRGLTRFNQATATATGRUSSDP--NLONIPVRPLGGRIRRA 598
QY 330 LKOPKSVYLVSPY-----FVPTKSGTDALAKLVODIDVTVLNS-----LOATDVA 377
DB 599 FIAEGWLLVALATYSOIELRYVAHLISGDBENLRIVQEGRDITHTETASMMFGVPRADV-- 656
QY 378 AVHSGVYKRPDLKA-----GIKYLEDPNNAVP----- 407
DB 657 -----PLMRAAKTINFGVLYGMSAHRLSQELAIPEEAQAFETRYFGSPRYR 705
QY 408 ---ATKDKGLTGSVTSIHAFTYVDGKRIFIGSFNLDPRSARLNTMGVIESP---K 460
DB 706 AMIENTLEEGRRRGIVETLF-----GRRRVDPLEARVKSVAERAAER-MAFNMPVOGT 757
QY 461 IAEQMERLTADSP---EYAVRVTLDNRHNLQWHDPAETRYPNPEAKMKRIAKLIS 517
DB 758 AADLKLKLMVKKLPRLLEGAMGAMLLQVHDELVLAPKERA---EAVARLAKVEWEGVYP 813
QY 518 L-LPIE 522
DB 814 LAVPLE 819

```

## RESULT 8

US-09-940-244-69  
 Sequence 69, Application US/09940244  
 Publication No. US20030044796A1

## GENERAL INFORMATION:

APPLICANT: Neri, Bruce P.  
 APPLICANT: Hall, Jeff G.  
 APPLICANT: Lyamichev, Victor  
 APPLICANT: Smith, Lloyd M.  
 TITLE OF INVENTION: Reactions on Dendrimers  
 FILE REFERENCE: FORS-06478



CURRENT APPLICATION NUMBER: US/09/940,244  
CURRENT FILING DATE: 2002-05-06  
NUMBER OF SEQ ID NOS: 422  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO: 69  
LENGTH: 833  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-940-244-69

Query Match 3.9%; Score 106; DB 9; Length 833;  
Best Local Similarity 21.8%; Pred. No. 1.2;  
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

QY 118 NLYVLAEEGVYR-----RLLDNNRGL-----DDLALDS-----HPNIEVRLFNPF 163  
DB 321 DILALAAAGGVHRAPEPYKALRDLKEARGLAKDLVLALEGLRPPGDDPML----- 376  
QY 164 VLKRRALGYLD---FPLNRHMNKSF---ADNRAT-----LGRNIGDE----- 206  
DB 377 -----LAYLDPSVTTEGVARRGGWTEAGERRALSERLPANLWGRLEGEERLLM 429  
QY 207 -YFKVGE--DYVFADDLIATG-----SVGEVS-----HDFDRYW 239  
DB 430 LYREVERPLSAVLAHME--ATGVRLDYAVLRALSLEVAGEIARLEAEVFRLAGHPN--- 484  
QY 240 ASHSANATRIIRSGNIGGLALGYNDTSRHA-----LTRYRETV 281  
DB 485 -LNSRDLERVLFD---ELGLPAIGTTEKTKRSTSAVLEALREAHPIVEKILQYRELT 540  
QY 282 EQ-----SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRRRKPIAGRLQDA 329  
DB 541 KIKSTYIDPLDPLIHRTGRHTFRNQATATGRSSSP--NLQNIPTVPLGQIRRA 598  
QY 330 LKQPEKSVYLVSPY-----FVPTKSGTDALAKLVODGIDVYLVTS-----LQATDVA 377  
DB 599 FLAEEGMLLVALDYQIELRVLAHLSGDENLIRVFOEGRIDHTEKASWMEGVPREAVD-- 656  
QY 378 AVHSGVYKRRPLKA-----GIKLYELQPNHVP----- 407  
DB 657 -----PLMRAAKTINFGVLYGMSAHRLSQELAIPEEAQAFTEIFYFQSPKVR 705  
QY 408 ---ATKDKGLTSSVTSLSHAKTFIYDGKRIFGSFNLDPRSARLNTENGVIESP---K 460  
DB 706 AMIEKTELEGRRRGVETLF-----GRRRYVPDLAEAVKSVREAER--MAFMNVPVGT 757  
QY 461 IAEOMERTIADSP---EYAYRVTLDRNRLQWHPDPTKTYPNPEEAKLMKRIAKILS 517  
DB 758 AADLKLAWVKLFRLPEMGARMLQVHDELVLAEKERA---EAVARLAKEVMEGVYP 813  
QY 518 L-LPIE 522  
DB 814 LAVPLE 819

RESULT 9  
US-09-940-925A-85  
Sequence 85, Application US/09940925A  
Publication No. US2003005438A1  
GENERAL INFORMATION:  
APPLICANT: BROW, MARY ANN D.  
LYAMICHEV, VICTOR I.  
OLIVE, DAVID M.  
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF  
PATHOGENS  
NUMBER OF SEQUENCES: 165  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MEDLEN & CARROLL  
STREET: 220 MONTGOMERY STREET, SUITE 2200  
CITY: SAN FRANCISCO  
STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/940,925A  
FILING DATE: 10-Jun-2002  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: CARROLL, PETER G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: FORS-01756  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 85:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 85:  
US-09-940-925A-85

Query Match 3.9%; Score 106; DB 9; Length 833;  
Best Local Similarity 21.8%; Pred. No. 1.2;  
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;

QY 118 NLYVLAEEGVYR-----RLLDNNRGL-----DDLALDS-----HPNIEVRLFNPF 163  
DB 321 DILALAAAGGVHRAPEPYKALRDLKEARGLAKDLVLALEGLRPPGDDPML----- 376  
QY 164 VLKRRALGYLD---FPLNRHMNKSF---ADNRAT-----LGRNIGDE----- 206  
DB 377 -----LAYLDPSVTTEGVARRGGWTEAGERRALSERLPANLWGRLEGEERLLM 429  
QY 207 -YFKVGE--DYVFADDLIATG-----SVGEVS-----HDFDRYW 239  
DB 430 LYREVERPLSAVLAHME--ATGVRLDYAVLRALSLEVAGEIARLEAEVFRLAGHPN--- 484  
QY 240 ASHSANATRIIRSGNIGGLALGYNDTSRHA-----LTRYRETV 281  
DB 485 -LNSRDLERVLFD---ELGLPAIGTTEKTKRSTSAVLEALREAHPIVEKILQYRELT 540  
QY 282 EQ-----SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRRRKPIAGRLQDA 329  
DB 541 KIKSTYIDPLDPLIHRTGRHTFRNQATATGRSSSP--NLQNIPTVPLGQIRRA 598  
QY 330 LKQPEKSVYLVSPY-----FVPTKSGTDALAKLVODGIDVYLVTS-----LQATDVA 377  
DB 599 FLAEEGMLLVALDYQIELRVLAHLSGDENLIRVFOEGRIDHTEKASWMEGVPREAVD-- 656  
QY 378 AVHSGVYKRRPLKA-----GIKLYELQPNHVP----- 407  
DB 657 -----PLMRAAKTINFGVLYGMSAHRLSQELAIPEEAQAFTEIFYFQSPKVR 705  
QY 408 ---ATKDKGLTSSVTSLSHAKTFIYDGKRIFGSFNLDPRSARLNTENGVIESP---K 460  
DB 706 AMIEKTELEGRRRGVETLF-----GRRRYVPDLAEAVKSVREAER--MAFMNVPVGT 757  
QY 461 IAEOMERTIADSP---EYAYRVTLDRNRLQWHPDPTKTYPNPEEAKLMKRIAKILS 517  
DB 758 AADLKLAWVKLFRLPEMGARMLQVHDELVLAEKERA---EAVARLAKEVMEGVYP 813  
QY 518 L-LPIE 522  
DB 814 LAVPLE 819

RESULT 10  
US-10-033-297-66  
Sequence 66, Application US/10033297  
Publication No. US20020187486A1  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
Lyamlichev, Victor I.  
Mast, Andrea L.  
Brow, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
Sequential Invasive Cleavages  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/033,297  
FILING DATE: 12-NOV-1996  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,597  
FILING DATE: 09-JUL-1999  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/756,386  
FILING DATE: 29-NOV-1996  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02736  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 66:  
US-10-033-297-66  
Query Match 3.9% Score 105; DB 9; Length 833;  
Best Local Similarity 21.5%; Pred. No. 1.5;  
Matches 116; Conservative 65; Mismatches 183; Indels 176; Gaps 27;  
QY 118 NLVYLAERGVV-----RLLDNNTRGL---DDLILADS-----HPNIEVRLNPF 163  
DB 321 DLALAAARGVHRAPEYKALRDLKEARGLADSLVALAREGLGLPQDDPML----- 376  
QY 164 VLKRRALGIYLD-----FPLRLRRHANKSFT--ADNRATI-----LGRNIGDE---- 206  
DB 377 -----LAYLDPSTNTPEGVARYRGEMTEAGERALSERLPLANLWGRIEGERILM 429

QY 207 -YFKGE-----DTVPADLILATGSSVYGENSHDFRYWAS-----HSAN 245  
DB 430 LYREVERPLSAVLAMHAEAGVRLDVAYLRLALSLEVAEELIALEAEVFLAGHPNLNSD 489  
QY 246 NATRIIRSGNICKGLQALQVNDETSNA-----LIRYREVED----- 283  
DB 490 QERVLFD--ELGLPAIKGTERGTGRSTSAVALEALREAHPIVEKILQRLTLKSTY 546  
QY 284 -SPLYOKI--QTGRIDMVSQVOT-----RLISDDPAKGLDRDRKPPPIAGRLDALKQPEK 335  
DB 547 IDPLPDLIHPRIGRILHTRNGTATATGRSSSDP--NLQNIIVRPFLGRIRARITAE 604  
QY 336 SYLVSPY-----PYPTKSGDALAKLVQDQIDVTYLTNS-----LQATDVAAVHSQY 383  
DB 605 WLVLADYSQIELRYLAHLSGDENLIRVQEGRDHTETASWMMFGVPREAVD----- 656  
QY 384 VKRRPLKA-----GIKLYELQPNHVP-----AT 409  
DB 657 -----PLMRRAKTIIEGVLVYKMSAHRLSQELALIFEEQATIERYFQSFPRVAMIEKT 711  
QY 410 KDKGLTGSSTSLHAKTFIVDGKRIEFGSFLNDPSPARLNTMGVIESP--KIAEQME 466  
DB 712 LIEGRRRGVETLF-----GRRRYVPLDEARVKSVSREAER--NAFNPVQGTADLMK 763  
QY 467 RLLADTSP---EYAYRVTLDRNNRLQWHDPRATRKTYPNPEKMKRRLAKILSL-LPIE 522  
DB 764 IAMVKLFPRLPEEMGARMILLQVHNEVLVLEAPKERA---EAVARLAKVEWGYPLAVPLE 819  
RESULT 11  
US-10-033-297-71  
Sequence 71, Application US/10033297  
Publication No. US20020187486A1  
GENERAL INFORMATION:  
APPLICANT: Hall, Jeff G.  
Lyamlichev, Victor I.  
Mast, Andrea L.  
Brow, Mary Ann D.  
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
Sequential Invasive Cleavages  
NUMBER OF SEQUENCES: 163  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/033,297  
FILING DATE: 12-NOV-1996  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,597  
FILING DATE: 09-JUL-1999  
APPLICATION NUMBER: US/08/823,516  
FILING DATE: 24-MAR-1997  
APPLICATION NUMBER: PCT/US97/01072  
FILING DATE: 21-JAN-1997  
APPLICATION NUMBER: US 08/759,038  
FILING DATE: 02-DEC-1996  
APPLICATION NUMBER: US 08/758,314  
FILING DATE: 29-NOV-1996  
APPLICATION NUMBER: US 08/682,853  
FILING DATE: 12-JUL-1996  
APPLICATION NUMBER: US 08/599,491

;; FILING DATE: 24-JAN-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Ingolia, Diane E.  
;; REGISTRATION NUMBER: 40,027  
;; REFERENCE/DOCKET NUMBER: FORS-02736  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (415) 705-8410  
;; TELEFAX: (415) 397-8338  
;; INFORMATION FOR SEQ ID NO: 71:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 833 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 71:  
US-10-033-297-71  
  
Query Match 3.9%; Score 105; DB 9; Length 833;  
Best local Similarity 21.8%; Pred. No. 1.5;  
Matches 119; Conservative 65; Mismatches 174; Indels 188; Gaps 30;  
  
QY 118 NLVYLAERGVY-----RLLDNNTRGL---DDLALDS---HPNIEVRLFNPF 163  
DB 321 DLALAAAGRGVHRAPEPYKALRDLEAKGLAKDLVLALEGLPGDDPML----- 376  
QY 164 VLKRWALGYLTD---FPLNRRMHKSFT--ADNRATI-----LGRNIGDE---- 206  
DB 377 -----LAYLDPSTNTPGVARRYGGEWTEBAGERALSERLFANLWGLGEEERLW 429  
QY 207 -YFKVGE--DTVPADLDIATG-----SVGGEVS-----HDFDRYW 239  
DB 430 LYREVERPLSAVLAHME--ATGVRLDVAYLRLSLLEVAGELRLAEVFLAGHPFN--- 484  
QY 240 ASHSANATRIIRSGNIGKGLALGYNDETSRA-----LLRYRETV 281  
DB 485 -LNSRQLERVLFD---ELGLPAIGKTEKGRKSTSAVLEALREAHPIVEKILQYREL 540  
QY 282 EQ-----SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRRKPPIAGLQDA 329  
DB 541 KIKSTYIDPLDIHPRGTGLHTRFNQGTATGRLSSDP--NLQNIPTPLGQRIIRRA 598  
QY 330 LKQPEKSVLVSPY-----FVPTKSGTDALAKLVODGIDVTVLTS-----LQATDVA 377  
DB 599 FLAESEWMLVALVYSQIELRVLAHLSGDENLIRVQEGRDHETETASMMGVPREAVD-- 656  
QY 378 AVHSGVYKTRKPLKA-----GIKLYELQPNHVP-----AT 409  
DB 657 -----PLMRAAKTINFGVLYGMSAHLSELAIPYEAQAFIERFYQSEPKVRA 705  
QY 408 ---ATKDKGLGSSVTSLSHAKTFIVDKRIFIGSFNLDPRSARLNTMGVLTSP---K 460  
DB 706 AWIEKLEGRRGVETLF-----GRRRYVPDEARVKSVEAER--MAFNMPVOGT 757  
QY 461 IAEQMERTLADTSP---EYAYRVTLDRHNRLOMHPATKRTYPNPEPAKIMKRIAKILS 517  
DB 758 AADLMKLAIVKLFPRLEEGARMLQVHDELVLPAEKERA---EAVARLAKEVMEGVYP 813  
QY 518 L-LPIE 522  
DB 814 LAVPLE 819  
  
RESULT 12  
US-09-940-244-66  
; Sequence 66, Application US/09940244  
; Publication No. US20030044796A1  
; GENERAL INFORMATION:  
; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamlichev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478

;; CURRENT APPLICATION NUMBER: US/09/940,244  
;; CURRENT FILING DATE: 2002-05-06  
;; NUMBER OF SEQ ID NOS: 422  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 66  
;; LENGTH: 833  
;; TYPE: PRP  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Synthetic  
US-09-940-244-66  
  
Query Match 3.9%; Score 105; DB 9; Length 833;  
Best local Similarity 21.5%; Pred. No. 1.5;  
Matches 116; Conservative 65; Mismatches 183; Indels 176; Gaps 27;  
  
QY 118 NLVYLAERGVY-----RLLDNNTRGL---DDLALDS---HPNIEVRLFNPF 163  
DB 321 DLALAAAGRGVHRAPEPYKALRDLEAKGLAKDLVLALEGLPGDDPML----- 376  
QY 164 VLKRWALGYLTD---FPLNRRMHKSFT--ADNRATI-----LGRNIGDE---- 206  
DB 377 -----LAYLDPSTNTPGVARRYGGEWTEBAGERALSERLFANLWGLGEEERLW 429  
QY 207 -YFKVGE-----DTVPADLDIATGSGVGEVSHDFRYWAS-----HSAH 245  
DB 430 LYREVERPLSAVLAHMEATGVRLDVAYLRLSLLEVAGELRLAEVFLAGHPFNLSRD 489  
QY 246 NATRIIRSGNIGKGLALGYNDETSRA-----LLRYRETVQ----- 283  
DB 490 QLREVRVLFDD---ELGLPAIGKTEKGRKSTSAVLEALREAHPIVEKILQYRELTKIKSTY 546  
QY 284 -SPLYOKI--QTGRIDMOSVOT-----RLISDDPAKGLDRRKPPIAGLQDALQPEK 335  
DB 547 IDPLDIHPRGTGLHTRFNQGTATGRLSSDP--NLQNIPTPLGQRIIRRA 604  
QY 336 SVLVSPY-----FVPTKSGTDALAKLVODGIDVTVLTS-----LQATDVAHSGY 383  
DB 605 WLIVVALDSQIELRVLAHLSGDENLIRVQEGRDHETETASMMGVPREAVD----- 656  
QY 384 VKYRKPLKA-----GIKLYELQPNHVP-----AT 409  
DB 657 -----PLMRAAKTINFGVLYGMSAHLSELAIPYEAQAFIERFYQSEPKVRA 711  
QY 410 KDKGLGSSVTSLSHAKTFIVDKRIFIGSFNLDPRSARLNTMGVLTSP---KIDOME 466  
DB 712 LEBGRRGVETLF-----GRRRYVPDEARVKSVEAER--MAFNMPVOGTAAADLMK 763  
QY 467 RTLADTSP---EYAYRVTLDRHNRLOMHPATKRTYPNPEPAKIMKRIAKILSL-LPIE 522  
DB 764 LAVVKKLFPRLEEGARMLQVHDELVLPAEKERA---EAVARLAKEVMEGVYPLAVPLE 819  
  
RESULT 13  
US-09-940-244-71  
; Sequence 71, Application US/09940244  
; Publication No. US20030044796A1  
; GENERAL INFORMATION:  
; APPLICANT: Neri, Bruce P.  
; APPLICANT: Hall, Jeff G.  
; APPLICANT: Lyamlichev, Victor  
; APPLICANT: Smith, Lloyd M.  
; TITLE OF INVENTION: Reactions on Dendrimers  
; FILE REFERENCE: FORS-06478  
; CURRENT APPLICATION NUMBER: US/09/940,244  
; CURRENT FILING DATE: 2002-05-06  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 71  
; LENGTH: 833  
; TYPE: PRP  
; ORGANISM: Artificial Sequence  
; FEATURE:

```

1  PRIOR FILING DATE: 2000-10-23
2  PRIOR APPLICATION NUMBER: 60/253,625
3  PRIOR FILING DATE: 2000-11-27
4  PRIOR APPLICATION NUMBER: 60/257,931
5  PRIOR FILING DATE: 2000-12-22
6  PRIOR APPLICATION NUMBER: 60/269,308
7  PRIOR FILING DATE: 2001-02-16
8  NUMBER OF SEQ ID NOS: 14110
9  SOFTWARE: FastSeq for Windows Version 4.0
10 SEQ ID NO 13285
11 LENGTH: 486
12 TYPE: PRT
13 ORGANISM: Streptococcus pneumoniae
14 US-09-815-242-13285
15
16 Query Match          3.8%; Score 103.5; DB 10; Length 486;
17 Best Local Similarity 19.0%; Pred. NO. 0.9;
18 Matches 77; Conservative 55; Mismatches 114; Indels 159; Gaps
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
98
```



;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087607  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; Remaining Prior Application data removed - See file wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 116  
;; SEQ ID NO 42  
;; LENGTH: 489  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-245-143-42

Query Match 3.8%; Score 102.5; DB 9; Length 489;  
Best Local Similarity 25.0%; Pred. No. 1.1;  
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

QY 284 SPLXQIQIQT-GTIDMOSVOTRLISDDPAKGLDRDRKRP---PIAGRLQDALQPEKSVYL 339  
DB 60 SPAMEPLEAEARQQRDSCQLVLESIP-QDLPSSAGSPSAQPLGQAWLQDLDTAQSVAHV 118  
QY 340 VSPYFVPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAY 379  
DB 119 ASYYSLSLGPDIQGVNDSSQGLQALLQKQLQGLGRNISLAVATSSPTLARTSTDLQVLA 178  
QY 380 HSGVYKRRKPL--LKAGIKLYELQPNHAVPAKDKGLGSSVTSLSHAKTFYDGKRIFIG 437  
DB 179 RGAHVR-QVPMGRLTRGV-----LHSKFVWVDGRHIYWG 211  
QY 438 SFNLDPRSARLNTMGVYV-ESPKIAEQMERT 468  
DB 212 SANMDMRSLTQYKELGAVIYNCSHLAQDLEKT 243

RESULT 18  
US-10-245-771-42  
;; Sequence 42, Application US/10245771  
;; Publication No. US20030068781A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Baker, Kevin  
;; APPLICANT: Ealon, Dan  
;; APPLICANT: Filvaroff, Ellen  
;; APPLICANT: Goddard, Audrey  
;; APPLICANT: Grimaldi, J. Christopher  
;; APPLICANT: Gurney, Austin  
;; APPLICANT: Smith, Victoria  
;; APPLICANT: Stephan, Jean-Phillippe  
;; APPLICANT: Watanabe, Colin  
;; APPLICANT: Wood, William  
;; APPLICANT: Zhang, Zemin  
;; APPLICANT: Fong, Sherman  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; FILE REFERENCE: P3630R1C98  
;; CURRENT APPLICATION NUMBER: US/10/245,771  
;; PRIOR APPLICATION NUMBER: 10/197942  
;; PRIOR FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/059114  
;; PRIOR FILING DATE: 1997-09-17  
;; PRIOR APPLICATION NUMBER: 60/063046  
;; PRIOR FILING DATE: 1997-10-24  
;; PRIOR APPLICATION NUMBER: 60/065027  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689

;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087607  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; Remaining Prior Application data removed - See file wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 116  
;; SEQ ID NO 42  
;; LENGTH: 489  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-245-771-42

Query Match 3.8%; Score 102.5; DB 9; Length 489;  
Best Local Similarity 25.0%; Pred. No. 1.1;  
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

QY 284 SPLXQIQIQT-GTIDMOSVOTRLISDDPAKGLDRDRKRP---PIAGRLQDALQPEKSVYL 339  
DB 60 SPAMEPLEAEARQQRDSCQLVLESIP-QDLPSSAGSPSAQPLGQAWLQDLDTAQSVAHV 118  
QY 340 VSPYFVPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAY 379  
DB 119 ASYYSLSLGPDIQGVNDSSQGLQALLQKQLQGLGRNISLAVATSSPTLARTSTDLQVLA 178  
QY 380 HSGVYKRRKPL--LKAGIKLYELQPNHAVPAKDKGLGSSVTSLSHAKTFYDGKRIFIG 437  
DB 179 RGAHVR-QVPMGRLTRGV-----LHSKFVWVDGRHIYWG 211  
QY 438 SFNLDPRSARLNTMGVYV-ESPKIAEQMERT 468  
DB 212 SANMDMRSLTQYKELGAVIYNCSHLAQDLEKT 243

RESULT 19  
US-10-245-851-42  
;; Sequence 42, Application US/10245851  
;; Publication No. US20030068782A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Baker, Kevin  
;; APPLICANT: Ealon, Dan  
;; APPLICANT: Filvaroff, Ellen  
;; APPLICANT: Goddard, Audrey  
;; APPLICANT: Grimaldi, J. Christopher  
;; APPLICANT: Gurney, Austin  
;; APPLICANT: Smith, Victoria  
;; APPLICANT: Stephan, Jean-Phillippe  
;; APPLICANT: Watanabe, Colin  
;; APPLICANT: Wood, William  
;; APPLICANT: Zhang, Zemin  
;; APPLICANT: Fong, Sherman  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; FILE REFERENCE: P3630R1C93  
;; CURRENT APPLICATION NUMBER: US/10/245,851  
;; PRIOR APPLICATION NUMBER: 10/197942  
;; PRIOR FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/059114  
;; PRIOR FILING DATE: 1997-09-17  
;; PRIOR APPLICATION NUMBER: 60/063046  
;; PRIOR FILING DATE: 1997-10-24  
;; PRIOR APPLICATION NUMBER: 60/065027  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478

```
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-851-42

Query Match          3.8%; Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

OY 284 SPLXOKIOT-GRIDMSVOTRLISDDPAKGLDRDRKP---PIAGRLQDALAKQPEKSYVL 339
DB 60 SPAMEPLEEAKQQRDSCQLVESIP-ODLPSAAGSPSAQPLGQAWLQDLDTAQSVAHV 118
OY 340 VSPYFVPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAY 379
DB 119 ASYVMSLTGPDIGVNDSSQGLGALLQKQLGRNISLAVATSSFTLARTSTDLQVLA 178
OY 380 HSGYVYRKRP--LKAGIKLYELQPNHAVPATRKGLTGSSVTSLAKTFIVDGKRIFTG 437
DB 179 RGAHYR-QVPMGRLTRGV-----LHSEKFWVDGRIIYMG 211
OY 438 SFNLDPRSARLNTMGVVI-ESPKIAEQMERT 468
DB 212 SANMDMRSLTQYKELGAVIYNCSHLAQDEKLT 243

RESULT 20
US-10-245-883-42
;; Sequence 42, Application US/10245883
;; Publication No. US20030068783A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C70
;; CURRENT APPLICATION NUMBER: US/10/245.883
;; PRIOR FILING DATE: 2002-09-16
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
```

```
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-883-42

Query Match          3.8%; Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

OY 284 SPLXOKIOT-GRIDMSVOTRLISDDPAKGLDRDRKP---PIAGRLQDALAKQPEKSYVL 339
DB 60 SPAMEPLEEAKQQRDSCQLVESIP-ODLPSAAGSPSAQPLGQAWLQDLDTAQSVAHV 118
OY 340 VSPYFVPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAY 379
DB 119 ASYVMSLTGPDIGVNDSSQGLGALLQKQLGRNISLAVATSSFTLARTSTDLQVLA 178
OY 380 HSGYVYRKRP--LKAGIKLYELQPNHAVPATRKGLTGSSVTSLAKTFIVDGKRIFTG 437
DB 179 RGAHYR-QVPMGRLTRGV-----LHSEKFWVDGRIIYMG 211
OY 438 SFNLDPRSARLNTMGVVI-ESPKIAEQMERT 468
DB 212 SANMDMRSLTQYKELGAVIYNCSHLAQDEKLT 243

RESULT 21
US-10-237-535-42
;; Sequence 42, Application US/10237535
;; Publication No. US20030073188A1
;; GENERAL INFORMATION:
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watanabe, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C3
;; CURRENT APPLICATION NUMBER: US/10/237.535
;; PRIOR FILING DATE: 2002-09-06
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
```

;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/091358  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/099803  
;; PRIOR FILING DATE: 1998-09-10  
;; PRIOR APPLICATION NUMBER: 60/106932  
;; PRIOR FILING DATE: 1998-11-03  
;; PRIOR APPLICATION NUMBER: 60/115554  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/119342  
;; PRIOR FILING DATE: 1999-02-09  
;; PRIOR APPLICATION NUMBER: 60/123957  
;; PRIOR FILING DATE: 1999-03-12  
;; PRIOR APPLICATION NUMBER: 60/123972  
;; PRIOR FILING DATE: 1999-03-11  
;; PRIOR APPLICATION NUMBER: 60/127372  
;; PRIOR FILING DATE: 1999-04-01  
;; PRIOR APPLICATION NUMBER: 60/131271  
;; PRIOR FILING DATE: 1999-04-27  
;; PRIOR APPLICATION NUMBER: 60/133459  
;; PRIOR FILING DATE: 1999-05-11  
;; PRIOR APPLICATION NUMBER: 60/135725  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/135729  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/135750  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/138385  
;; PRIOR FILING DATE: 1999-06-09  
;; PRIOR APPLICATION NUMBER: 60/140653  
;; PRIOR FILING DATE: 1999-06-22  
;; PRIOR APPLICATION NUMBER: 60/141037  
;; PRIOR FILING DATE: 1999-06-23  
;; PRIOR APPLICATION NUMBER: 60/144732  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/144758  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/144790  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/145228  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/145698  
;; PRIOR FILING DATE: 1999-07-26  
;; PRIOR APPLICATION NUMBER: 60/146222  
;; PRIOR FILING DATE: 1999-07-28  
;; PRIOR APPLICATION NUMBER: 60/146843  
;; PRIOR FILING DATE: 1999-08-03  
;; PRIOR APPLICATION NUMBER: 60/148188  
;; PRIOR FILING DATE: 1999-08-10  
;; PRIOR APPLICATION NUMBER: 60/148513  
;; PRIOR FILING DATE: 1999-08-12  
;; PRIOR APPLICATION NUMBER: 60/149327  
;; PRIOR FILING DATE: 1999-08-17  
;; PRIOR APPLICATION NUMBER: 60/149395  
;; PRIOR FILING DATE: 1999-08-17  
;; PRIOR APPLICATION NUMBER: 60/150114  
;; PRIOR FILING DATE: 1999-08-20  
;; PRIOR APPLICATION NUMBER: 60/151700  
;; PRIOR FILING DATE: 1999-08-31  
;; PRIOR APPLICATION NUMBER: 60/151734  
;; PRIOR FILING DATE: 1999-08-31  
;; PRIOR APPLICATION NUMBER: 60/162506  
;; PRIOR FILING DATE: 1999-10-29  
;; PRIOR APPLICATION NUMBER: 60/170262  
;; PRIOR FILING DATE: 1999-12-09  
;; PRIOR APPLICATION NUMBER: 60/177118  
;; PRIOR FILING DATE: 2000-01-20  
;; PRIOR APPLICATION NUMBER: 60/179851  
;; PRIOR FILING DATE: 2000-02-02  
;; PRIOR APPLICATION NUMBER: 60/180921  
;; PRIOR FILING DATE: 2000-02-08  
;; PRIOR APPLICATION NUMBER: 60/187202  
;; PRIOR FILING DATE: 2000-03-03  
;; PRIOR APPLICATION NUMBER: 60/198587  
;; PRIOR FILING DATE: 2000-04-18  
;; PRIOR APPLICATION NUMBER: 60/199614  
;; PRIOR FILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: 60/206330  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/206368  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/209832  
;; PRIOR FILING DATE: 2000-06-05  
;; PRIOR APPLICATION NUMBER: 60/218371  
;; PRIOR FILING DATE: 2000-07-13  
;; PRIOR APPLICATION NUMBER: 60/222695  
;; PRIOR FILING DATE: 2000-08-02  
;; PRIOR APPLICATION NUMBER: 60/229896  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/230621  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/232887  
;; PRIOR FILING DATE: 2000-09-15  
;; PRIOR APPLICATION NUMBER: 60/235147  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/261878  
;; PRIOR FILING DATE: 2001-01-12  
;; PRIOR APPLICATION NUMBER: 60/261910  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/261939  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/262150  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/264395  
;; PRIOR FILING DATE: 2001-01-25  
;; PRIOR APPLICATION NUMBER: 60/266421  
;; PRIOR FILING DATE: 2001-02-02  
;; PRIOR APPLICATION NUMBER: 60/267623  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/274399  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 60/280982  
;; PRIOR FILING DATE: 2001-04-03  
;; PRIOR APPLICATION NUMBER: 60/282129  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/282199  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/290589  
;; PRIOR FILING DATE: 2001-05-09  
;; PRIOR APPLICATION NUMBER: 09/180997  
;; PRIOR FILING DATE: 1998-11-19  
;; PRIOR APPLICATION NUMBER: 09/267213  
;; PRIOR FILING DATE: 1999-03-12  
;; PRIOR APPLICATION NUMBER: 09/380137  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 09/380138  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 09/403297  
;; PRIOR FILING DATE: 1999-10-18  
;; PRIOR APPLICATION NUMBER: 09/423741  
;; PRIOR FILING DATE: 1999-11-10  
;; PRIOR APPLICATION NUMBER: 09/709238  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 09/802706  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 09/872035  
;; PRIOR FILING DATE: 2001-06-01  
;; PRIOR APPLICATION NUMBER: 09/918585  
;; PRIOR FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: 09/924419



PRIOR FILING DATE: 1998-03-27

PRIOR	APPLICATION	NUMBER	60/0986478
PRIOR	FILING DATE	1998-05-22	
PRIOR	APPLICATION	NUMBER	60/087607
PRIOR	FILING DATE	1998-06-02	
PRIOR	APPLICATION	NUMBER	60/098801
PRIOR	FILING DATE	1998-06-18	
PRIOR	APPLICATION	NUMBER	60/0905517
PRIOR	FILING DATE	1998-06-24	
PRIOR	APPLICATION	NUMBER	60/090688
PRIOR	FILING DATE	1998-06-25	
PRIOR	APPLICATION	NUMBER	60/091356
PRIOR	FILING DATE	1998-07-01	
PRIOR	APPLICATION	NUMBER	60/091917
PRIOR	FILING DATE	1998-07-07	
PRIOR	APPLICATION	NUMBER	60/099801
PRIOR	FILING DATE	1998-09-10	
PRIOR	APPLICATION	NUMBER	60/106933
PRIOR	FILING DATE	1998-11-03	
PRIOR	APPLICATION	NUMBER	60/115555
PRIOR	FILING DATE	1999-01-12	
PRIOR	APPLICATION	NUMBER	60/119344
PRIOR	FILING DATE	1999-02-09	
PRIOR	APPLICATION	NUMBER	60/123955
PRIOR	FILING DATE	1999-03-12	
PRIOR	APPLICATION	NUMBER	60/123977
PRIOR	FILING DATE	1999-03-11	
PRIOR	APPLICATION	NUMBER	60/127377
PRIOR	FILING DATE	1999-04-01	
PRIOR	APPLICATION	NUMBER	60/131271
PRIOR	FILING DATE	1999-04-27	
PRIOR	APPLICATION	NUMBER	60/134555
PRIOR	FILING DATE	1999-05-11	
PRIOR	APPLICATION	NUMBER	60/135722
PRIOR	FILING DATE	1999-05-25	
PRIOR	APPLICATION	NUMBER	60/135722
PRIOR	FILING DATE	1999-05-25	
PRIOR	APPLICATION	NUMBER	60/135755
PRIOR	FILING DATE	1999-05-25	
PRIOR	APPLICATION	NUMBER	60/138388
PRIOR	FILING DATE	1999-06-09	
PRIOR	APPLICATION	NUMBER	60/140655
PRIOR	FILING DATE	1999-06-22	
PRIOR	APPLICATION	NUMBER	60/141033
PRIOR	FILING DATE	1999-06-23	
PRIOR	APPLICATION	NUMBER	60/144733
PRIOR	FILING DATE	1999-07-20	
PRIOR	APPLICATION	NUMBER	60/145222
PRIOR	FILING DATE	1999-07-20	
PRIOR	APPLICATION	NUMBER	60/145699
PRIOR	FILING DATE	1999-07-26	
PRIOR	APPLICATION	NUMBER	60/146222
PRIOR	FILING DATE	1999-07-28	
PRIOR	APPLICATION	NUMBER	60/146844
PRIOR	FILING DATE	1999-08-03	
PRIOR	APPLICATION	NUMBER	60/148188
PRIOR	FILING DATE	1999-08-10	
PRIOR	APPLICATION	NUMBER	60/148511
PRIOR	FILING DATE	1999-08-12	
PRIOR	APPLICATION	NUMBER	60/149322
PRIOR	FILING DATE	1999-08-17	
PRIOR	APPLICATION	NUMBER	60/149399
PRIOR	FILING DATE	1999-08-17	
PRIOR	APPLICATION	NUMBER	60/150111
PRIOR	FILING DATE	1999-08-20	
PRIOR	APPLICATION	NUMBER	60/151700
PRIOR	FILING DATE	1999-08-31	
PRIOR	APPLICATION	NUMBER	60/151733
PRIOR	FILING DATE	1999-08-31	
PRIOR	APPLICATION	NUMBER	60/162506

;; PRIOR FILING DATE: 1999-10-29  
;; PRIOR APPLICATION NUMBER: 60/170262  
;; PRIOR FILING DATE: 1999-12-09  
;; PRIOR APPLICATION NUMBER: 60/177118  
;; PRIOR FILING DATE: 2000-01-20  
;; PRIOR APPLICATION NUMBER: 60/179851  
;; PRIOR FILING DATE: 2000-02-02  
;; PRIOR APPLICATION NUMBER: 60/180921  
;; PRIOR FILING DATE: 2000-02-08  
;; PRIOR APPLICATION NUMBER: 60/187202  
;; PRIOR FILING DATE: 2000-03-03  
;; PRIOR APPLICATION NUMBER: 60/198587  
;; PRIOR FILING DATE: 2000-04-18  
;; PRIOR APPLICATION NUMBER: 60/199614  
;; PRIOR FILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: 60/206330  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/206368  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/209832  
;; PRIOR FILING DATE: 2000-06-05  
;; PRIOR APPLICATION NUMBER: 60/218371  
;; PRIOR FILING DATE: 2000-07-13  
;; PRIOR APPLICATION NUMBER: 60/222695  
;; PRIOR FILING DATE: 2000-08-02  
;; PRIOR APPLICATION NUMBER: 60/229896  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/230621  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/232887  
;; PRIOR FILING DATE: 2000-09-15  
;; PRIOR APPLICATION NUMBER: 60/235147  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: 60/261878  
;; PRIOR FILING DATE: 2001-01-12  
;; PRIOR APPLICATION NUMBER: 60/261910  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/261939  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/262150  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/264395  
;; PRIOR FILING DATE: 2001-01-25  
;; PRIOR APPLICATION NUMBER: 60/266421  
;; PRIOR FILING DATE: 2001-02-02  
;; PRIOR APPLICATION NUMBER: 60/267623  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/274399  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 60/280982  
;; PRIOR FILING DATE: 2001-04-03  
;; PRIOR APPLICATION NUMBER: 60/282129  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/282199  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/290589  
;; PRIOR FILING DATE: 2001-05-09  
;; PRIOR APPLICATION NUMBER: 60/180997  
;; PRIOR FILING DATE: 1998-11-19  
;; PRIOR APPLICATION NUMBER: 60/267213  
;; PRIOR FILING DATE: 1999-03-12  
;; PRIOR APPLICATION NUMBER: 60/380137  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 60/380138  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 60/403297  
;; PRIOR FILING DATE: 1999-10-18  
;; PRIOR APPLICATION NUMBER: 60/423741  
;; PRIOR FILING DATE: 1999-11-10  
;; PRIOR APPLICATION NUMBER: 60/709238  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/802706  
;; PRIOR FILING DATE: 2001-03-09

;; PRIOR APPLICATION NUMBER: 09/872035  
;; PRIOR FILING DATE: 2001-06-01  
;; PRIOR APPLICATION NUMBER: 09/918585  
;; PRIOR FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: 09/924419  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: 09/927796  
;; PRIOR FILING DATE: 2001-08-09  
;; PRIOR APPLICATION NUMBER: 09/929404  
;; PRIOR FILING DATE: 2001-08-13  
;; PRIOR APPLICATION NUMBER: 09/931836  
;; PRIOR FILING DATE: 2001-08-16  
;; PRIOR APPLICATION NUMBER: 09/941992  
;; PRIOR FILING DATE: 2001-08-28  
;; PRIOR APPLICATION NUMBER: 09/946374  
;; PRIOR FILING DATE: 2001-09-04  
;; PRIOR APPLICATION NUMBER: 10/001054  
;; PRIOR FILING DATE: 2001-11-30  
;; PRIOR APPLICATION NUMBER: 10/052586  
;; PRIOR FILING DATE: 2002-01-15  
;; PRIOR APPLICATION NUMBER: 10/081056  
;; PRIOR FILING DATE: 2002-02-20  
;; PRIOR APPLICATION NUMBER: 10/119480  
;; PRIOR FILING DATE: 2002-04-09

Query Match 3.8%; Score 102.5; DB 9; Length 489;  
Best Local Similarity 25.0%; Pred. No. 1.1;  
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

QY 284 SPYKRIQT-GRIDQSVTRILSDPAKGLDRDRKP---PIAGRLQALQKPEKSVYL 339  
DB 60 SPAMEPLEAEARQQRDCQLLVESIP-QDLPSAAGSPSAPGLQAWLDLDTAGSVAIV 118  
QY 340 VSEFYVPT-----KSGTDLAKLVQ---DGIDVTVLNS---LQATD--VAAV 379  
DB 119 ASYWSLTGPDIGVNDSSQLGEALLQKQLLGRNISIATAVTSPTLARTSTDLQVLA 178  
QY 380 HSGYVYRKPL--LKGITLYELQPNHVAIPATRKDKLTSSVTSLSHAKFTYDGRKFIYG 437  
DB 179 RGAHVR-QVPMGRLTRGV-----LHSEKRVVVDGRITVYG 211

QY 438 SFNLDPRSARLNTMGVVI-ESPKIAEOMERT 468  
DB 212 SANMBWRSLTQYKELGAVIYNCSHLAQLDEKT 243

RESULT 23  
US-10-238-283-42  
; Sequence 42, Application US/10238283  
; Publication No. US20030073190A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Eaton, Dan  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Philippe  
; APPLICANT: Watanabe, Colin  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Fong, Sherman  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3630R1C15  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/238, 283  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/059114  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/063046



```
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-055-42
```

```
Query Match          3.8%; Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;
```

```
OY 284 SPLYOKIQT-GRIDMSQVTRLISDDPAKGLDRDRKP---PIAGRLQDALKQPEKSYVL 339
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 60 SPAMEPLEAEARQQRDSCQLVESIP-QDLPASAGSFAQPLGQAWLQLDLTAQESVHV 118
OY 340 VSPYEVPT-----KSGTDALARKVQ---DGIDVTYLTNS---LQATD--VAAY 379
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 119 ASYWSLPGDIDGVNDSSQJGEALLQKIQQLGRNISLAVATSSPTLARTSTDQVLAA 178
OY 380 HSGYKRYKRL--LKAQIKLYELQPNHAVPATKDKGLGSSVTSJHAKTFYDGRIFIG 437
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 179 RGAHVR-QVPMGRRLRGV-----LHSEFWVVDGHHIYWG 211
OY 438 SFNLDPRSARLWTEMGVVI-ESPKIAEQMERT 468
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 212 SANMDWRSLTOYKELGAVIYNCSHLAQDLEKT 243
```

```
RESULT 26
US-10-245-147-42
; Sequence 42, Application US/10245147
; Publication No. US20030073193A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C72
; CURRENT APPLICATION NUMBER: US/10/245,147
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
```

```
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-147-42
```

```
Query Match          3.8%; Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;
```

```
OY 284 SPLYOKIQT-GRIDMSQVTRLISDDPAKGLDRDRKP---PIAGRLQDALKQPEKSYVL 339
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 60 SPAMEPLEAEARQQRDSCQLVESIP-QDLPASAGSFAQPLGQAWLQLDLTAQESVHV 118
OY 340 VSPYEVPT-----KSGTDALARKVQ---DGIDVTYLTNS---LQATD--VAAY 379
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 119 ASYWSLPGDIDGVNDSSQJGEALLQKIQQLGRNISLAVATSSPTLARTSTDQVLAA 178
OY 380 HSGYKRYKRL--LKAQIKLYELQPNHAVPATKDKGLGSSVTSJHAKTFYDGRIFIG 437
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 179 RGAHVR-QVPMGRRLRGV-----LHSEFWVVDGHHIYWG 211
OY 438 SFNLDPRSARLWTEMGVVI-ESPKIAEQMERT 468
| | : : : : | | : : : : | | : : : : | | : : : : | | : : : : |
DB 212 SANMDWRSLTOYKELGAVIYNCSHLAQDLEKT 243
```

```
RESULT 27
US-10-245-730-42
; Sequence 42, Application US/10245730
; Publication No. US20030073194A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watande, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C85
; CURRENT APPLICATION NUMBER: US/10/245,730
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
```

```

PRIORITY FILLING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-06-25
Remaining PRIOR Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 42
LENGTH: 489
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-739-42

Query Match      3.8% Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches    53; Conservative   33; Mismatches    71; Indels     55; Gaps    11.

OY      284 SPLVQKIQTF-CRIMWQSVOFRLISDDPAKKGLDRDRRKP---PIAGRLDALDKPEKSIVYL 339
         ||::||::| |::||::| |::||::| |::||::| |::||::| |::||::| |::||::| |::||::|
Db       60 SPAMEPLEAEARQQNDSCQLVLVESIP-QDPLEASAGSPSACPLGGAQMILDTLTADESVAHV 118
OY      340 VSPFEVP-----KGSTDLAKLVQ---DGIDVTLYTNS---LQATD--VAAV 379
         ::::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db       119 ASYWLSLTGPDIGVDSSSQELALQKLQQLGRNISLAIVATSSPTLARSTDQLVLAA 178
OY      380 HSGYVKRKRP--LKAGIKELYEQNHNAVPRPKDKGLGSSTSLHAKFIYDGKRIFIG 437
         ::::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db       179 RGAHVR-QVPMGRLLRGV-----LHSKFVVVDGNRIIYG 211
OY      438 SFNLDPARSARLNTEGVVI-ESEPKIAEOMERT 468
         I::II::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db       212 SANMDWRSLTVQEKEIGAIVYMCSHAADLEKT 243

RESULT 29
US-10-246-210-42
Sequence 42, Application US/10246210
Publication No. US20030073196A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Phillippe
APPLICANT: Watanabe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fongy, Sherman
TITLE OR INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630RIC121
CURRENT APPLICATION NUMBER: US/10/246, 210
CURRENT FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: 10/197942
PRIOR FILING DATE: 2002-07-18
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090557
```

```

; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-246-210-42

Query Match
Best Local Similarity 25.0%; Pred. No. 1.1; Length 489;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

QY 284 SPLYKIQIQT-GRIDQSVOTRLISDDPAKGLDRDRRP---PIAGRLDALKQPKESVYL 339
DB 60 SPAMELEAEARQQRSCQLVLESIP-QDLPSAAGSPAQGLGQAWLQDLDTAQESVHV 118
QY 340 VSPYEPYPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAV 379
DB 119 ASYWSLTGPDPDGVNDSSQLGEALLQKLOLLGRNISLAVATSSPTLARTSTDLQVLA 178
QY 380 HSGYVKRKPL--LKAGIKLYELOPNHAYPATRKDGLTSSVTSIHAKEFIYDGRKIFIG 437
DB 179 RGAHVR-QVPMGRRLTRGV-----LHSEFWVVDGRHITWG 211
QY 438 SFNLDPRSARLNTMGVVI-ESPKIAEQMERT 468
DB 212 SANMDWRSILTQVKEGAVIYNCSHLAQDEKT 243

RESULT 30
US-10-239-196-42
; Sequence 42, Application US/10239196
; Publication No. US20030074735A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C8
; CURRENT FILING DATE: 2002-09-09
; PRIOR FILING DATE: 10/197942
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR FILING DATE: 1997-10-24
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
```

```

; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-239-196-42

Query Match
Best Local Similarity 25.0%; Pred. No. 1.1; Length 489;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

QY 284 SPLYKIQIQT-GRIDQSVOTRLISDDPAKGLDRDRRP---PIAGRLDALKQPKESVYL 339
DB 60 SPAMELEAEARQQRSCQLVLESIP-QDLPSAAGSPAQGLGQAWLQDLDTAQESVHV 118
QY 340 VSPYEPYPT-----KSGTDALAKLVQ---DGIDVTVLNS---LQATD--VAAV 379
DB 119 ASYWSLTGPDPDGVNDSSQLGEALLQKLOLLGRNISLAVATSSPTLARTSTDLQVLA 178
QY 380 HSGYVKRKPL--LKAGIKLYELOPNHAYPATRKDGLTSSVTSIHAKEFIYDGRKIFIG 437
DB 179 RGAHVR-QVPMGRRLTRGV-----LHSEFWVVDGRHITWG 211
QY 438 SFNLDPRSARLNTMGVVI-ESPKIAEQMERT 468
DB 212 SANMDWRSILTQVKEGAVIYNCSHLAQDEKT 243

RESULT 31
US-10-243-024-42
; Sequence 42, Application US/10243024
; Publication No. US2003007741A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C38
; CURRENT FILING DATE: 2002-09-12
; PRIOR FILING DATE: 10/197942
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR FILING DATE: 1997-10-24
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```

; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-243-024-42

```

Query Match	3.8%;	Score 102.5;	DB 9;	Length 489;
Best Local Similarity	25.0%;	Pred. No. 1.1;		
Matches 53; Conservative	33;	Mismatches 71;	Indels 55;	Gaps 11

```

OY 284 SPLXQIOT-GRIDMOSVOTRLISDDPAKGLDRORRRP---PIAGRIQDALKQEKSVYL 333
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
Db 60 SPANPELEAEARQORDSCQLVLYSIP-QDLPASAGSTSAOPLQAMQLDLOTRAQESVHV 118
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 340 VSPYEVPT-----KSGTDAALKVYO---DGIDVYVLTSN---LQATD--VAAY 379
    || : : : : || : : : : || : : : : || : : : : || : : : : || : : : :
Db 119 ASYWSLTLGPDIGVNDSSQJGEALLKQLQGLGRNLSLAVATSSPLTARSTDLQYLAA 178
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 380 HSGYVKKRPV--LKAGIKELYELOPNNAVPATKDKGLTSSWSVSLAKTFIVDGKRIFFIG 433
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 RGAHVH-OVPMGRILTRGV-----LHSEKFWVDGRHTYMG 211
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 438 SFNLDPSRARSINTEMGVYI-ESPKIAEQMERT 468
    || : : || : : || : : || : : || : : || : : || : : || : : || : :
Db 212 SANDMWSLTQVKKELGAVITVCSHLADOLEKT 243

```

```

RESULT 32
US-10-243-409-42
; Sequence 42, Application US/10243409
; Publication No. US2003007742A1
; GENERAL INFORMATION:

```

APPLICANT: Baker, Kevin  
 APPLICANT: Baton, Dan  
 APPLICANT: Elivaroff, Ellen  
 APPLICANT: Goddard, Audrey  
 APPLICANT: Grimaldi, J. Christopher  
 APPLICANT: Gurney, Austin  
 APPLICANT: Smith, Victoria  
 APPLICANT: Stephan, Jean-Phillippe  
 APPLICANT: Watande, Colin  
 APPLICANT: Wood, William  
 APPLICANT: Zhang, Zemin  
 APPLICANT: Fong, Sherman  
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 FILE REFERENCE: P3630R1C41

```

; Remaining Prior Application data removed - See File Wrapper or PALM
; NUMBER OF SEQ ID NOS: 116 .
; SEQ ID NO 42

```

```

.; LENGTH: 489
.; TYPE: PRT
.; ORGANISM: Homo Sapien
US-10-243-409-42

```

Query Match	3.8%	Score 102.5;	DB 9;	Length 489;
Best Local Similarity	25.0%;	Pred. No. 1.1;		
Matches	53;	Conservative	33;	Mismatches 71;
				Indels 55;
				Gaps 11

[illegible]

RESULT 33  
US-10-245-033-42  
; Sequence 42, Application US/10245033  
; Publication No. US20030078401A1

```

: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Eaton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goodard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Phillippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3630R1C84
: CURRENT APPLICATION NUMBER: US/10/245,033

```

```

; SEQ ID NO 42
;
; LENGTH: 489
;
; TYPE: prt
;

```





Best Local Similarity 25.0%; Pred. No. 1:1;  
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

```

QY  ,284 SPLOKIOQT-GRIDMOSVOIRLLISDDPAKGLDRNRKRP---PIAGRIQLDALKQPEKSVYL 3359
    ||:::|||||
Db  60 SPAPELEAEARQQRDSCQLVLESTP-QDLPSPAAGSPSAQPLGQAWQLLDPAQESVHV 118
    ||:::|||||
QY  340 VSPFVFPVPT-----KSGFDALAKLVQ---DGIDVVLVINS-----LQATD--VAAV 379
    ||:::|||||
Db  119 ASYYWSLGTGPDICGVNDSQSOLGEMLLQKLOQLGRNLSLAVATSSPLTATSTIDVQVLA 178
    ||:::|||||
QY  380 HSGVVKYRKPL--LKAGIKLYELOPNHNAVPA TKDKGLTGSSVYLSLAKTFIVDKRIFIG 437
    ||:::|||||
Db  179 RGAHVH-QVPMGRILTRGV-----LHSGFMVVDGRHITYMG 211
    ||:::|||||
QY  438 SFNIDPRSARLNTMGVYI--ESPKIAQEMERT 468
    ||:::|||||
Db  212 SANDMRSLTOVQKELGAVIYINCSHLADLEKT 243
    ||:::|||||

```

RESULT 36  
US-10-243

```

Sequence 42: Application US/10243095
Publication No. US20030082726A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Eaton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Matanbe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Rong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEOTIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3630R1C43
CURRENT APPLICATION NUMBER: US/10/243,095
CURRENT FILING DATE: 2002-09-12
PRIORITY APPLICATION NUMBER: 10/197942
PRIORITY FILING DATE: 2002-07-18
PRIORITY APPLICATION NUMBER: 60/059114
PRIORITY FILING DATE: 1997-09-17
PRIORITY APPLICATION NUMBER: 60/063046
PRIORITY FILING DATE: 1997-10-24
PRIORITY APPLICATION NUMBER: 60/065027
PRIORITY FILING DATE: 1997-11-10
PRIORITY APPLICATION NUMBER: 60/079669
PRIORITY FILING DATE: 1998-03-27
PRIORITY APPLICATION NUMBER: 60/086478
PRIORITY FILING DATE: 1998-05-22
PRIORITY APPLICATION NUMBER: 60/087607
PRIORITY FILING DATE: 1998-06-02
PRIORITY APPLICATION NUMBER: 60/089801
PRIORITY FILING DATE: 1998-06-18
PRIORITY APPLICATION NUMBER: 60/090557
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090669
PRIORITY FILING DATE: 1998-06-25
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 42
LENGTH: 489
TYPE: PRT
ORGANISM: Homo Sapien
US-10-243-095-42

```

Query Match	3.8%;	Score 102.5;	DB 9;	Length 489;
Best Local Similarity	25.0%;	Pred. No. 1.1;		
Matches	53;	Conservative	33;	Mismatches 71;
				Indels 55;
				Gaps 11;

```

QY 284 SFLQKQKQKT-GRIDMOSVQTRLLISDDPAKGDPRRRKPR---PLAGRLQDMLKQPEKSVYL 339
Db 60 SPAPELEAEARKQQRDSCQLVLVPSIP-QDLPISAAGSPSAQPIGQAWQLDITRAQSEVHY 118
QY 340 VSPFVPEVPT-----KSGTALAKLVQ--DGIDVTVLTNS-----LQATD--VAAV 379
Db 119 ASYWSLITGPDIGVNDSSSQGEALHOKLQQLGRLNSLAVATSSPTLATSTIDLYLAA 178
QY 380 HSGVVKYKRRP--LKAGIKLYELQPNHAVPATKDKGLTGSSVYTSLAKTFYVDGKRIFIG 437
Db 179 RGAHVR-QVPMGRILTRGV-----LHSGFMVVDGRHYMG 211
QY 438 SFNDDPSARLNTEMGYVI--ESPFLAQEMERT 468
Db 212 SANDDWRSLTQVKELGAVIYNCSHLADDELT 243

```

RESULT 37  
US-10-245

Sequence 42, Application US/10245185  
Publication No. US20030082728A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin  
APPLICANT: Falon, Dan  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Goddard, Audrey  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Guney, Austin  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Philippe  
APPLICANT: Watanabe, Colin  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
APPLICANT: Fong Sherman  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3630R1C80  
CURRENT APPLICATION NUMBER: US/10/245,185  
CURRENT FILING DATE: 2002-09-16  
PRIORITY APPLICATION NUMBER: 10/197942  
PRIORITY FILING DATE: 2002-07-18  
PRIORITY APPLICATION NUMBER: 60/059114  
PRIORITY FILING DATE: 1997-09-17  
PRIORITY APPLICATION NUMBER: 60/063046  
PRIORITY FILING DATE: 1997-10-24  
PRIORITY APPLICATION NUMBER: 60/065027  
PRIORITY FILING DATE: 1997-11-10  
PRIORITY APPLICATION NUMBER: 60/079689  
PRIORITY FILING DATE: 1998-03-27  
PRIORITY APPLICATION NUMBER: 60/086478  
PRIORITY FILING DATE: 1998-05-22  
PRIORITY APPLICATION NUMBER: 60/087607  
PRIORITY FILING DATE: 1998-06-02  
PRIORITY APPLICATION NUMBER: 60/089801  
PRIORITY FILING DATE: 1998-06-18  
PRIORITY APPLICATION NUMBER: 60/090557  
PRIORITY FILING DATE: 1998-06-24  
PRIORITY APPLICATION NUMBER: 60/090669  
PRIORITY FILING DATE: 1998-06-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 42  
LENGTH: 489  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-245-185-42

Query Match	3.8%	Score 102.5;	DB 9;	Length 489;
Best Local Similarity	25.0%	Pred. No. 1.1;		
Matches	53;	Conservative	33;	Mismatches 71;
				Indels 55;
				Gaps 11.
284	SPLYKQIT-GRIMQSVOTRLISDDPAKGLDRDRRK---	DIAGRLDALKQPEKSYVL	339	

```
Db 60 SPAMEPLEAARQDRSCQVIVESIP--QDLPASAGSPSAQPLGQAWMLQDLDTAQESVHV 118
Oy 340 VSPFVPT-----KSGTDALAKLVQ---DCIDVTVLTNS-----LQATD--VAAV 379
Db 119 ASYWSLTPGPDIGVNDSSQLGEALLQKQLQGLGRNISLAVNTSSPTLARTSTDLQVLA 178
Oy 380 HSGYVYKRPPL--LKAGITLLEYLQPNHAVPATRKCLTSSVTSLSAKTFIYDGKRIFTG 437
Db 179 RGAHVR-QVPMGRLTRGV-----LHSEKFWVDGRRHYMG 211
Oy 438 SFNLDPRSARLNTMGVVI--ESPRIAEOMERT 468
Db 212 SANMDRSLTYQKELGAVIYNCSHLAQDEKT 243

RESULT 38
US-10-245-427-42
: Sequence 42, Application US/10245427
: Publication No. US20030082729A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Eaton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Philippe
: APPLICANT: Watanabe, Collin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630RIC68
: CURRENT APPLICATION NUMBER: US/10/245,427
: PRIOR FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: 10/197942
: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: PRIOR FILING DATE: 1998-06-25
: Remaining Prior Application data removed - See file Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 42
: LENGTH: 489
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-245-427-42

Query Match 3.8%, Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

Oy 284 SPLVYKIQT-GRIDWQSVQTRILISDDPAKGLDRDRRP---PIAGRLQDALAKQPEKSVYL 339
Db 60 SPAMEPLEAARQDRSCQVIVESIP--QDLPASAGSPSAQPLGQAWMLQDLDTAQESVHV 118
```

```
Oy 340 VSPFVPT-----KSGTDALAKLVQ---DCIDVTVLTNS-----LQATD--VAAV 379
Db 119 ASYWSLTPGPDIGVNDSSQLGEALLQKQLQGLGRNISLAVNTSSPTLARTSTDLQVLA 178
Oy 380 HSGYVYKRPPL--LKAGITLLEYLQPNHAVPATRKCLTSSVTSLSAKTFIYDGKRIFTG 437
Db 179 RGAHVR-QVPMGRLTRGV-----LHSEKFWVDGRRHYMG 211
Oy 438 SFNLDPRSARLNTMGVVI--ESPRIAEOMERT 468
Db 212 SANMDRSLTYQKELGAVIYNCSHLAQDEKT 243

RESULT 39
US-10-245-473-42
: Sequence 42, Application US/10245473
: Publication No. US20030082730A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Eaton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Philippe
: APPLICANT: Watanabe, Collin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630RIC110
: CURRENT APPLICATION NUMBER: US/10/245,473
: PRIOR FILING DATE: 2002-09-17
: PRIOR APPLICATION NUMBER: 10/197942
: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: PRIOR FILING DATE: 1998-06-25
: Remaining Prior Application data removed - See file Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 42
: LENGTH: 489
: TYPE: PRT
: ORGANISM: Homo Sapien
US-10-245-473-42

Query Match 3.8%, Score 102.5; DB 9; Length 489;
Best Local Similarity 25.0%; Pred. No. 1.1;
Matches 53; Conservative 33; Mismatches 71; Indels 55; Gaps 11;

Oy 284 SPLVYKIQT-GRIDWQSVQTRILISDDPAKGLDRDRRP---PIAGRLQDALAKQPEKSVYL 339
Db 60 SPAMEPLEAARQDRSCQVIVESIP--QDLPASAGSPSAQPLGQAWMLQDLDTAQESVHV 118
Oy 340 VSPFVPT-----KSGTDALAKLVQ---DCIDVTVLTNS-----LQATD--VAAV 379
```

Db 119 ASYWSLTGPDIGVNDSSSQGEALLQKLOQLGRNLSIAVAIASSPTLARTSDIDQLAA 178  
 QY 380 HSGVYKTRKPL--LKAGIKLEYLOPNHNAVPATKDKGLTSSVYTSLAKFTIVDGKRIFIG 437  
 Db 179 RGAVHR-QVPMGRILTRGV-----LHSEFVVDVGRIHYMG 211  
 QY 438 SFNLDPRSARLNTFMGVII--ESPRIAEQMENT 468  
 Db 212 SANDMRSLTQVKEGLGAVITNCISLAQDLEKT 243

```

RESULT 40
US-10-245-770-42
Sequence 42: Application US/10245770
Publication NO. US2003082731A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin
APPLICANT: Baton, Dan
APPLICANT: Filvaroff, Ellen
APPLICANT: Goddard, Audrey
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin
APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe
APPLICANT: Matanbe, Colin
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Rong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C74
CURRENT APPLICATION NUMBER: US/10/245,770
CURRENT FILING DATE: 2002-09-16
PRIORITY APPLICATION NUMBER: 10/197942
PRIORITY FILING DATE: 2002-07-18
PRIORITY APPLICATION NUMBER: 60/059114
PRIORITY FILING DATE: 1997-09-17
PRIORITY APPLICATION NUMBER: 60/063046
PRIORITY FILING DATE: 1997-10-24
PRIORITY APPLICATION NUMBER: 60/065027
PRIORITY FILING DATE: 1997-11-10
PRIORITY APPLICATION NUMBER: 60/079689
PRIORITY FILING DATE: 1998-03-27
PRIORITY APPLICATION NUMBER: 60/086478
PRIORITY FILING DATE: 1998-05-22
PRIORITY APPLICATION NUMBER: 60/087607
PRIORITY FILING DATE: 1998-06-02
PRIORITY APPLICATION NUMBER: 60/089801
PRIORITY FILING DATE: 1998-06-18
PRIORITY APPLICATION NUMBER: 60/090557
PRIORITY FILING DATE: 1998-06-24
PRIORITY APPLICATION NUMBER: 60/090669
PRIORITY FILING DATE: 1998-06-25
Remaining SEQ ID Application data removed - See File Wrapper or PAM.
NUMBER OF SEQ ID NOS: 116
SEQ ID NO 42
LENGTH: 489
TYPE: PRT
ORGANISM: Homo Sapien
US-10-245-770-42

```

Query Match	3.8%;	Score 102.5;	DB 9;	Length 489;
Best Local Similarity	25.0%;	Pred. No. 1.1;		
Matches	53;	Conservative	33;	Mismatches 71;
				Indels 55;
				Gaps 11;

OY	284	SPLTKLQIOT-CRIMOSVOTFLISDDPAKGIIDRRRP	---	PIAGRQDALKQEPKSVYL	339
		: : :		: : :	
Db	60	SPANEPLLEAKRQQRDCQLVVESTI-QDLPSAAGSP	SPMQPLGQAWHLQDLTRQESVHV	118	
OY	340	VSFPFVPI-----KSGTDALKLVO-----	DGIDVTVLNS-----LOATD--VAAV	379	
		: : :	: : :	: : :	
Db	119	ASYWSLUGPDIGVNDSSSQGEHLQDLQDLGRNLSL	VAWSSPTLARTSDLOVLA	178	

```
OY 380 HSGVKKKPKP- -LKAGIKLTELPHNAVPA TKKGJLTSSVTSLAKTETIVGKRFTIG 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 RGAHVR-QVPNGRILRGV-----LHSKEFVVDSGRHTIYG 211
```

```
Search completed: May 12, 2003, 10:06:14
Job time : 62 secs
```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: May 12, 2003, 10:06:22 ; Search time 16 Seconds  
(without alignments)  
965.440 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 555

Sequence: 1 MRANPKTQAMPSETISLMT.....KMKRIAKILSLPIESL 525

Scoring table: OLIGO

Searched: 262574 seqs, 29422922 residues

Word size: 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCrus.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.3	18	2	US-08-480-190-173
2	7	1.3	18	2	US-08-488-379-173
3	7	1.3	18	5	PCT-US93-07545-173
4	7	1.3	19	2	US-08-480-190-172
5	7	1.3	19	2	US-08-488-379-172
6	7	1.3	19	5	PCT-US93-07545-172
7	7	1.3	21	1	US-07-768-269-2
8	7	1.3	21	1	US-07-768-269-43
9	7	1.3	21	1	US-07-768-269-56
10	7	1.3	21	1	US-07-837-780-5
11	7	1.3	21	1	US-07-837-780-25
12	7	1.3	21	1	US-08-243-728-2
13	7	1.3	21	1	US-08-231-449-2
14	7	1.3	21	1	US-08-673-269-4
15	7	1.3	21	1	US-08-530-326-2
16	7	1.3	21	2	US-08-530-326-9
17	7	1.3	21	4	US-08-530-342A-2
18	7	1.3	21	4	US-08-530-342A-9
19	7	1.3	21	4	US-08-561-096-18
20	7	1.3	21	4	US-08-635-928-18
21	7	1.3	21	4	US-08-635-928-25
22	7	1.3	21	4	US-07-728-169-4
23	7	1.3	21	5	PCT-US93-03795-2
24	7	1.3	21	6	5231166-1
25	7	1.3	26	2	US-08-480-190-214
26	7	1.3	26	2	US-08-488-379-214
27	7	1.3	26	5	PCT-US93-07545-214

28	7	1.3	37	1	US-08-673-269-7
29	7	1.3	37	3	US-09-192-048-19
30	7	1.3	37	4	US-07-728-169-7
31	7	1.3	37	6	5231166-3
32	7	1.3	38	1	US-08-673-269-8
33	7	1.3	38	4	US-07-728-169-8
34	7	1.3	178	1	US-08-673-269-2
35	7	1.3	178	4	US-07-728-169-2
36	7	1.3	192	4	US-08-960-190A-38
37	7	1.3	229	3	US-08-470-535-12
38	7	1.3	238	4	US-09-149-476-485
39	7	1.3	246	3	US-09-248-335-32
40	7	1.3	248	1	US-08-644-664B-27
41	7	1.3	248	2	US-08-761-277A-27
42	7	1.3	252	4	US-09-134-001C-5519
43	7	1.3	253	2	US-08-484-905-109
44	7	1.3	253	3	US-08-481-985B-109
45	7	1.3	253	4	US-08-370-476-109
46	7	1.3	344	3	US-09-110-116-4
47	7	1.3	394	4	US-09-812-484-33
48	7	1.3	496	3	US-08-889-841B-16
49	7	1.3	513	4	US-08-924-183-8
50	7	1.3	513	4	US-09-488-364-8
51	7	1.3	521	4	US-08-956-322-4
52	7	1.3	601	1	US-08-458-477A-2
53	7	1.3	601	2	US-09-033-153-2
54	7	1.3	601	4	US-09-325-430B-2
55	7	1.3	652	3	US-09-110-116-1
56	7	1.3	652	4	US-08-956-322-2
57	7	1.3	656	4	US-08-765-081-5
58	7	1.3	696	3	US-09-098-082-5
59	7	1.3	696	5	PCT-US93-06994-5
60	7	1.3	703	5	PCT-US95-06994-8
61	7	1.3	1021	1	US-08-497-025-3
62	7	1.3	2071	4	US-09-415-522-6
63	7	1.1	9	2	US-08-845-926-39
64	6	1.1	9	4	US-09-518-046-35
65	6	1.1	9	4	US-09-518-046-129
66	6	1.1	10	2	US-09-351-296-39
67	6	1.1	10	2	US-08-617-929-39
68	6	1.1	13	4	US-09-258-754-105
69	6	1.1	13	4	US-09-042-107-105
70	6	1.1	14	2	US-08-617-929-41
71	6	1.1	14	2	US-08-392-806A-10
72	6	1.1	14	4	US-09-257-450-10
73	6	1.1	19	1	US-08-244-116B-6
74	6	1.1	19	3	US-08-848-580-8
75	6	1.1	19	5	PCT-US92-04537-4
76	6	1.1	20	1	US-08-273-669-7
77	6	1.1	20	2	US-08-617-929-6
78	6	1.1	20	2	US-08-954-724-7
79	6	1.1	20	2	US-08-966-947-7
80	6	1.1	21	1	US-08-243-728-9
81	6	1.1	21	1	US-08-243-728-10
82	6	1.1	21	3	US-08-848-580-9
83	6	1.1	21	5	PCT-US92-04537-5
84	6	1.1	21	6	5196511-11
85	6	1.1	22	2	US-08-146-028-94
86	6	1.1	22	2	US-08-146-028-95
87	6	1.1	22	2	US-08-146-028-151
88	6	1.1	22	4	US-08-723-425A-94
89	6	1.1	22	4	US-08-723-425A-95
90	6	1.1	22	4	US-08-723-425A-151
91	6	1.1	22	4	US-09-112-206-94
92	6	1.1	22	4	US-09-112-206-95
93	6	1.1	22	4	US-09-112-206-151
94	6	1.1	23	3	US-07-927-391-7
95	6	1.1	24	1	US-08-066-309-11
96	6	1.1	24	5	PCT-US94-05660-11
97	6	1.1	26	1	US-08-486-421-9
98	6	1.1	26	1	US-08-470-911-9
99	6	1.1	26	2	US-08-486-809-9
100	6	1.1	28	3	US-08-848-580-10

Sequence 7, Appli
Sequence 19, Appli
Sequence 7, Appli
Patent No. 5231166
Sequence 8, Appli
Sequence 2, Appli
Sequence 38, Appli
Sequence 12, Appli
Sequence 485, App
Sequence 32, Appli
Sequence 27, Appli
Sequence 5519, Ap
Sequence 109, App
Sequence 109, App
Sequence 4, Appli
Sequence 16, Appli
Sequence 8, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 2, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 8, Appli
Sequence 6, Appli
Sequence 39, Appli
Sequence 105, App
Sequence 10, Appli
Sequence 41, Appli
Sequence 10, Appli
Sequence 8, Appli
Sequence 7, Appli
Sequence 7, Appli
Sequence 6, Appli
Sequence 7, Appli
Sequence 9, Appli
Sequence 10, Appli
Sequence 9, Appli
Sequence 5, Appli
Patent No. 5196511
Sequence 94, Appli
Sequence 95, Appli
Sequence 151, App
Sequence 95, Appli
Sequence 151, App
Sequence 95, Appli
Sequence 11, Appli
Sequence 11, Appli
Sequence 9, Appli
Sequence 9, Appli
Sequence 10, Appli















977 5 1.0 16 2 US-08-660-592-7 Sequence 7, Appli  
978 5 1.0 16 2 US-08-660-592-8 Sequence 8, Appli  
979 5 1.0 16 2 US-08-660-592-11 Sequence 11, Appli  
980 5 1.0 16 2 US-08-928-615-5 Sequence 5, Appli  
981 5 1.0 16 2 US-08-676-242-12 Sequence 12, Appli  
982 5 1.0 16 4 US-08-602-999A-167 Sequence 167, App  
983 5 1.0 16 4 US-08-602-999A-220 Sequence 220, App  
984 5 1.0 16 4 US-08-602-999A-235 Sequence 235, App  
985 5 1.0 16 4 US-08-602-999A-365 Sequence 365, App  
986 5 1.0 16 4 US-09-166-930A-2 Sequence 2, Appli  
987 5 1.0 16 4 US-09-166-930A-3 Sequence 3, Appli  
988 5 1.0 16 4 US-09-166-930A-8 Sequence 8, Appli  
989 5 1.0 16 4 US-08-907-739-142 Sequence 142, App  
990 5 1.0 16 4 US-09-166-448-5 Sequence 5, Appli  
991 5 1.0 16 4 US-09-348-933-5 Sequence 5, Appli  
992 5 1.0 16 4 US-09-009-953-274 Sequence 274, App  
993 5 1.0 16 4 US-09-697-884-5 Sequence 5, Appli  
994 5 1.0 16 4 US-09-500-124-167 Sequence 167, App  
995 5 1.0 16 4 US-09-500-124-220 Sequence 220, App  
996 5 1.0 16 4 US-09-500-124-235 Sequence 235, App  
997 5 1.0 16 4 US-09-500-124-365 Sequence 365, App  
998 5 1.0 16 4 US-09-729-597-142 Sequence 142, App  
999 5 1.0 17 1 US-07-798-776-22 Sequence 22, Appli  
1000 5 1.0 17 1 US-08-061-062A-2 Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-480-190-173  
; Sequence 173, Application US/08480190  
; Patent No. 5827516  
; GENERAL INFORMATION:  
; APPLICANT: Robert G. Urban  
; APPLICANT: Roman M. Chicz  
; APPLICANT: Dario A. A. Vignali  
; APPLICANT: Mary L. Hedley  
; APPLICANT: Lawrence J. Stern  
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
; NUMBER OF SEQUENCES: 274  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,190  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/077,255  
; FILING DATE: June 15, 1993  
; APPLICATION NUMBER: 07/925,460  
; FILING DATE: August 11, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00246/168001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 173:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 18  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-480-190-173

Query Match 1.38; Score 7; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 364 VTVLINS 370  
Db 4 VTVLINS 10

RESULT 2  
US-08-488-379-173  
; Sequence 173, Application US/08488379  
; Patent No. 5880103  
; GENERAL INFORMATION:  
; APPLICANT: Robert G. Urban  
; APPLICANT: Roman M. Chicz  
; APPLICANT: Dario A. A. Vignali  
; APPLICANT: Mary L. Hedley  
; APPLICANT: Lawrence J. Stern  
; TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
; NUMBER OF SEQUENCES: 274  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: Wordperfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/488,379  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/077,255  
; FILING DATE: June 15, 1993  
; APPLICATION NUMBER: 07/925,460  
; FILING DATE: August 11, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00246/168001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 173:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 18  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-488-379-173

Query Match 1.38; Score 7; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 364 VTVLINS 370  
Db 4 VTVLINS 10

RESULT 3  
PCT-US93-07545-173  
Sequence 173, Application PC/TUS9307545  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07545  
FILING DATE: 19930811  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 173:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US93-07545-173

Query Match 1.3%; Score 7; DB 5; Length 18;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 364 VTVLINS 370  
|111111|  
Db 4 VTVLINS 10

RESULT 4  
US-08-480-190-172  
Sequence 172, Application US/08480190  
Patent No. 5627516  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 274  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street

CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,190  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,255  
FILING DATE: June 15, 1993  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 172:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-480-190-172

Query Match 1.3%; Score 7; DB 2; Length 19;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 364 VTVLINS 370  
|111111|  
Db 4 VTVLINS 10

RESULT 5  
US-08-488-379-172  
Sequence 172, Application US/08488379  
Patent No. 5880103  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 274  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,379  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/077,255  
FILING DATE: June 15, 1993  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 172:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-488-379-172

Query Match 1.3%; Score 7; DB 2; Length 19;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
Db 4 VTVLINS 10

RESULT 6  
PCT-US93-07545-172  
Sequence 172, Application PC/TUS9307545  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chiciz  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07545  
FILING DATE: 19930811  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 172:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 19  
TYPE: amino acid  
STRANDEDNESS:

TOPOLOGY: linear  
PCT-US93-07545-172

Query Match 1.3%; Score 7; DB 5; Length 19;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
Db 4 VTVLINS 10

RESULT 7  
US-07-768-269-2  
Sequence 2, Application US/07768269  
Patent No. 5306808  
GENERAL INFORMATION:  
APPLICANT: WAKIMASU, Mitsuhiro  
APPLICANT: KIKUCHI, Takashi  
APPLICANT: KUBO, Kazuki  
TITLE OF INVENTION: PEPTIDE DERIVATIVES AND PRODUCTION  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/768,269  
FILING DATE: 19911017  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 40762  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 SPRE UR

SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-768-269-2

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 8  
US-07-768-269-43  
Sequence 43, Application US/07768269  
Patent No. 5306808  
GENERAL INFORMATION:  
APPLICANT: WAKIMASU, Mitsuhiro  
APPLICANT: KIKUCHI, Takashi  
APPLICANT: KUBO, Kazuki

TITLE OF INVENTION: PEPTIDE DERIVATIVES AND PRODUCTION  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/768,269  
FILING DATE: 19911017  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 40762  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-768-269-43

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
|||||  
Db 1 CSCSSWL 7

RESULT 9  
US-07-768-269-56  
Sequence 56, Application US/07768269  
Patent No. 5306808  
GENERAL INFORMATION:  
APPLICANT: WAKIMASU, Mitsuhiro  
APPLICANT: KIKUCHI, Takashi  
APPLICANT: KUBO, Kazuki  
TITLE OF INVENTION: PEPTIDE DERIVATIVES AND PRODUCTION  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/768,269  
FILING DATE: 19911017  
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, David S  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 40762  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-768-269-56

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
|||||  
Db 1 CSCSSWL 7

RESULT 10  
US-07-837-780-5  
Sequence 5, Application US/07837780  
Patent No. 5352659  
GENERAL INFORMATION:  
APPLICANT: WAKIMASU, Mitsuhiro  
APPLICANT: KIKUCHI, Takashi  
APPLICANT: KUBO, Kazuki  
TITLE OF INVENTION: PEPTIDES, PRODUCTION AND USE THEREOF  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/837,780  
FILING DATE: 19920214  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: DAVID, Conlin G  
REGISTRATION NUMBER: 27026  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 1,15  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 3,11  
US-07-837-780-5

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
DB 1 CSCSSWL 7

RESULT 11  
US-07-837-780-25  
; Sequence 25, Application US/07837780  
; Patent No. 5352659  
; GENERAL INFORMATION:  
; APPLICANT: WAKIMASU, Mitsuhiro  
; APPLICANT: KIKUCHI, Takashi  
; APPLICANT: KUBO, Kazuki  
; TITLE OF INVENTION: PEPTIDES, PRODUCTION AND USE THEREOF  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/837,780  
; FILING DATE: 19920214  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DAVID, Conlin G  
; REGISTRATION NUMBER: 27026  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)523-3400  
; TELEFAX: (617)523-6440  
; TELEX: 200291 STRE UR  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FEATURE:  
; NAME/KEY: Disulfide-bond  
; LOCATION: 1,15  
; FEATURE:  
; NAME/KEY: Disulfide-bond  
; LOCATION: 3,11  
US-07-837-780-25

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
DB 1 CSCSSWL 7

RESULT 12  
US-08-243-728-2  
; Sequence 2, Application US/08243728  
; Patent No. 5492892  
; GENERAL INFORMATION:  
; APPLICANT: Andersen, Thomas T  
; APPLICANT: Spinnella, Michael J

TITLE OF INVENTION: ENDOTHELIN ANTAGONISTS  
NUMBER OF SEQUENCES: 84  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Scully, Scott, Murphy & Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: New York  
COUNTRY: United States  
ZIP: 11530

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/243,728  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 666,294  
FILING DATE: March 8, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: McNulty, William E.  
REGISTRATION NUMBER: 22,606  
REFERENCE/DOCKET NUMBER: 8104  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 742-4343  
TELEFAX: (516) 742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: both  
US-08-243-728-2

Query Match 1.3%; Score 7; DB 1; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
DB 1 CSCSSWL 7

RESULT 13  
US-08-231-449-2  
; Sequence 2, Application US/08231449  
; Patent No. 5616684  
; GENERAL INFORMATION:  
; APPLICANT: WAKIMASU, Mitsuhiro  
; APPLICANT: KIKUCHI, Takashi  
; APPLICANT: KAWADA, Akira  
; APPLICANT: SHIRAHUJI, Hideo  
; TITLE OF INVENTION: CYCLIC PEPTIDES AND USE THEREOF  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
; ADDRESSEE: CUSHMAN  
; STREET: 130 Water Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,449  
; FILING DATE:  
; CLASSIFICATION: 514



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/927,205  
FILING DATE: 07-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: NEUNER, George W  
REGISTRATION NUMBER: 26964  
REFERENCE/DOCKET NUMBER: 42016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-231-449-2

Query Match 1.3%; Score 7; DB 1;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 14  
US-08-673-269-4  
Sequence 4, Application US/08673269  
Patent No. 5756344  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OHKUBO, Shoichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
TITLE OF INVENTION: PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,269  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728169  
FILING DATE: 10-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-673-269-4

Query Match 1.3%; Score 7; DB 1;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 15  
US-08-530-326-2  
Sequence 2, Application US/08530326  
Patent No. 5961954  
GENERAL INFORMATION:  
APPLICANT: HILGER, CHRISTOPH-STEPH  
APPLICANT: DINKELBOG, LUDGER  
APPLICANT: KRAMP, WOLFGANG  
APPLICANT: SCHIER, HANS-MARTIN  
TITLE OF INVENTION: CHELATORS OF TYPE XN1S101 FOR RADIOACTIVE ISOTOPES, TH  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEISLER, DUBB, MEYER & LOVEJOY  
STREET: Four Embarcadero Center, Suite 400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,326  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE94/00372  
FILING DATE: 29-MAR-1994  
APPLICATION NUMBER: P 43 10 999.3  
FILING DATE: 31-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Dubb, Hubert E  
REGISTRATION NUMBER: 25,329  
REFERENCE/DOCKET NUMBER: WBLT 2006 (WO) PA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-362-3800  
TELEFAX: 415-362-2928  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-08-530-326-2

Query Match 1.3%; Score 7; DB 2;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 16  
US-08-530-326-9  
; Sequence 9, Application US/08530326  
; Patent No. 5961954  
; GENERAL INFORMATION:  
; APPLICANT: HILGER, CHRISTOPH-STEPH  
; APPLICANT: DINKELBORG, LUDGER  
; APPLICANT: KRAMP, WOLFGANG  
; APPLICANT: SCHIER, HANS-MARTIN  
; TITLE OF INVENTION: CHELATORS OF TYPE XN1S101 FOR RADIOACTIVE ISOTOPES, TH  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FLEISLER, DUBB, MEYER & LOVEJOY  
; STREET: Four Embarcadero Center, Suite 400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,326  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DE94/00372  
; FILING DATE: 29-MAR-1994  
; APPLICATION NUMBER: P 43 10 999.3  
; FILING DATE: 31-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dubb, Hubert E  
; REGISTRATION NUMBER: 25,329  
; REFERENCE/DOCKET NUMBER: WBLT 2006 (WO) PA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-362-3800  
; TELEFAX: 415-362-2928  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; US-08-530-326-9  
Query Match 1.3%; Score 7; DB 2; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7  
RESULT 17  
US-08-530-342A-2  
; Sequence 2, Application US/08530342A  
; Patent No. 6143275  
; GENERAL INFORMATION:  
; APPLICANT: HILGER, CHRISTOPH-STEPH  
; APPLICANT: DINKELBORG, LUDGER  
; APPLICANT: KRAMP, WOLFGANG  
; APPLICANT: SCHIER, HANS-MARTIN  
; TITLE OF INVENTION: TYPE S3N2 CHELATORS FOR RADIO  
; TITLE OF INVENTION: TYPE S3N2 CHELATORS FOR METAL COMPLEXES AND  
; TITLE OF INVENTION: ACTIVE ISOTOPES, THEIR METAL COMPLEXES AND

; TITLE OF INVENTION: THEIR DIAGNOSTIC AND THERAPEUTICAL USE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOOD, PHILLIPS, VAN SANTEN, CLARK &  
; ADDRESS: MORTIMER  
; STREET: 500 West Madison Street, Suite 3800  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60661-2511  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: WordPerfect for Windows 6.0a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/530,342A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/DE94/00370  
; FILING DATE: 29-MAR-1994  
; APPLICATION NUMBER: P 43 11 022.3  
; FILING DATE: 31-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mortimer, John S.  
; REGISTRATION NUMBER: 30,407  
; REFERENCE/DOCKET NUMBER: 1214, 00008  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/876-1800  
; TELEFAX: 312/876-2020  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 21 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; US-08-530-342A-2  
Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7  
RESULT 18  
US-08-530-342A-9  
; Sequence 9, Application US/08530342A  
; Patent No. 6143275  
; GENERAL INFORMATION:  
; APPLICANT: HILGER, CHRISTOPH-STEPH  
; APPLICANT: DINKELBORG, LUDGER  
; APPLICANT: KRAMP, WOLFGANG  
; APPLICANT: SCHIER, HANS-MARTIN  
; TITLE OF INVENTION: TYPE S3N2 CHELATORS FOR RADIO  
; TITLE OF INVENTION: ACTIVE ISOTOPES, THEIR METAL COMPLEXES AND  
; TITLE OF INVENTION: THEIR DIAGNOSTIC AND THERAPEUTICAL USE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOOD, PHILLIPS, VAN SANTEN, CLARK &  
; ADDRESS: MORTIMER  
; STREET: 500 West Madison Street, Suite 3800  
; CITY: Chicago  
; STATE: IL

COUNTRY: USA  
ZIP: 60661-2511  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: COMPAQ PROlinea 4/25S  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect for Windows 6.0a  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/530,342A  
FILING DATE: 08-MAR-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE94/00370  
FILING DATE: 29-MAR-1994  
APPLICATION NUMBER: P 43 11 022.3  
FILING DATE: 31-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mortimer, John S.  
REGISTRATION NUMBER: 30,407  
REFERENCE/DOCKET NUMBER: 1214, 00008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/876-1800  
TELEFAX: 312/876-2020  
TELEX:  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-08-530-342A-9

Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 19  
US-08-561-096-18  
Sequence 18, Application US/08561096  
Patent No. 6147051  
GENERAL INFORMATION:  
APPLICANT: WATANABE, Toshifumi  
APPLICANT: WAKIMASU, Mitsuhiro  
APPLICANT: KITADA, Chieko  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE PROPHYLACTIC  
TITLE OF INVENTION: AND/OR THERAPEUTIC TREATMENT OF ORGAN  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, REBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/561,096

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/231,017  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: NEUNER, George W  
REGISTRATION NUMBER: 26964  
REFERENCE/DOCKET NUMBER: 44349  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
STREET: 200291 STRE UR  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 1.15  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 3.11  
US-08-561-096-18

Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 20  
US-08-635-928-18  
Sequence 18, Application US/08635928  
Patent No. 6291639  
GENERAL INFORMATION:  
APPLICANT: CONRAD, JURGEN  
APPLICANT: DINKELBOG, LUDGER  
APPLICANT: ERBER, SEBASTIAN  
APPLICANT: FROMMELT, CORNELIUS  
APPLICANT: HOFME, WOLFGANG  
APPLICANT: KRAMP, WOLFGANG  
APPLICANT: KUTNER, GABRIELE  
APPLICANT: MALIN, REINHARD  
APPLICANT: SCHIER, HANS MARTIN  
APPLICANT: SCHNEIDER-MERGEN, JENS  
TITLE OF INVENTION: METAL-BINDING CYSTEINE-FREE  
TITLE OF INVENTION: PEPTIDES FOR DIAGNOSTIC AND  
TITLE OF INVENTION: THERAPEUTICAL PURPOSES.  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEISLER, DUBB, MEYER & LOVEJOY LLP  
STREET: Four Embarcadero Center, Suite 400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,928  
FILING DATE: 09/20/96  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DE94/01302  
FILING DATE: 10/27/94  
APPLICATION NUMBER: P4337599.5  
FILING DATE: 11/01/93  
ATTORNEY/AGENT INFORMATION:  
NAME: Molano, Michael A.  
REGISTRATION NUMBER: 39,777  
REFERENCE/DOCKET NUMBER: WBLT2010(WO/US)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-362-3800  
TELEFAX: 415-362-2928  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-08-635-928-18

Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 21  
US-08-635-928-25  
Sequence 25, Application US/08635928  
Patent No. 6291639  
GENERAL INFORMATION:  
APPLICANT: CONRAD, JURGEN  
APPLICANT: DINKELBORG, LUDGER  
APPLICANT: ERBER, SEBASTIAN  
APPLICANT: FROMMEL, CORNEIUS  
APPLICANT: HONNE, WOLFGANG  
APPLICANT: KRAMP, WOLFGANG  
APPLICANT: KUTNER, GABRIELE  
APPLICANT: MALIN, REINHARD  
APPLICANT: SCHIER, HANS MARTIN  
APPLICANT: SCHNEIDER-MERGENR, JENS  
TITLE OF INVENTION: METAL-BINDING CYSTEIN-FREE  
TITLE OF INVENTION: PEPTIDES FOR DIAGNOSTIC AND  
TITLE OF INVENTION: THERAPEUTICAL PURPOSES,  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEISLER, DUBB, MEYER & LOVEJOY LLP  
STREET: Four Embarcadero Center, Suite 400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 7.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/635,928  
FILING DATE: 09/20/96  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/DE94/01302  
FILING DATE: 10/27/94  
APPLICATION NUMBER: P4337599.5

FILING DATE: 11/01/93  
ATTORNEY/AGENT INFORMATION:  
NAME: Molano, Michael A.  
REGISTRATION NUMBER: 39,777  
REFERENCE/DOCKET NUMBER: WBLT2010(WO/US)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-362-3800  
TELEFAX: 415-362-2928  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: no  
ANTI-SENSE: no  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
US-08-635-928-25

Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 1 CSCSSWL 7

RESULT 22  
US-07-728-169-4  
Sequence 4, Application US/07728169  
Patent No. 6365725  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OKUBO, Shoichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
TITLE OF INVENTION: PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/728,169  
FILING DATE: 19910710  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-07-728-169-4

Query Match 1.3%; Score 7; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
|||||||  
DB 1 CSCSSWL 7

RESULT 23  
PCT-US93-03795-2  
Sequence 2, Application PC/TUS9303795

GENERAL INFORMATION:  
APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
TITLE OF INVENTION: METHODS FOR TREATING VASCULAR DISORDERS  
TITLE OF INVENTION: BY INHIBITING THE ENDOTHELIN STIMULATORY  
TITLE OF INVENTION: ACTIVITY OF TGFB  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: CAMPBELL & FLORES  
STREET: 4370 LA JOLLA VILLAGE DRIVE, SUITE 700  
CITY: SAN DIEGO  
STATE: CALIFORNIA  
COUNTRY: UNITED STATES  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03795  
FILING DATE: 19930421  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KONSRI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: PP-UC 9607  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-535-9001  
TELEFAX: 619-535-8949  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 21 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
PCT-US93-03795-2

Query Match 1.3%; Score 7; DB 5; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
|||||||  
DB 1 CSCSSWL 7

RESULT 24  
5231166-1

PATENT NO. 5231166  
APPLICANT: MASAKI, TOMOH;YAMAGISAWA, MASASHI;INOUE, AKIHITO  
TITLE OF INVENTION: ENDOTHELIN  
NUMBER OF SEQUENCES: 15  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/422,132  
FILING DATE: 16-OCT-1989  
SEQ ID NO: 1:  
LENGTH: 21  
5231166-1

Query Match 1.3%; Score 7; DB 6; Length 21;  
Best Local Similarity 100.0%; Pred. No. 8.2;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
|||||||  
DB 1 CSCSSWL 7

RESULT 25  
US-08-480-190-214  
Sequence 214, Application US/08480190

PATENT NO. 5827516  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz  
APPLICANT: Darío A. A. Vignall  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 274  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/480,190  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,255  
FILING DATE: June 15, 1993  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 214:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-480-190-214

Query Match 1.3%; Score 7; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
|||||||  
DB 13 VTVLINS 19

RESULT 26  
US-08-488-379-214  
Sequence 214, Application US/08488379

PATENT NO. 5880103  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz

APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 274  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,379  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/077,255  
FILING DATE: June 15, 1993  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 214:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-488-379-214

Query Match 1.3%; Score 7; DB 2; Length 26;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
1111111  
Db 13 VTVLINS 19

RESULT 27  
PCT-US93-07545-214  
Sequence 214, Application PC/TUS9307545  
GENERAL INFORMATION:  
APPLICANT: Robert G. Urban  
APPLICANT: Roman M. Chicz  
APPLICANT: Dario A. A. Vignali  
APPLICANT: Mary L. Hedley  
APPLICANT: Lawrence J. Stern  
APPLICANT: Jack L. Strominger  
TITLE OF INVENTION: IMMUNOMODULATORY PEPTIDES  
NUMBER OF SEQUENCES: 273  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: Wordperfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/07545  
FILING DATE: 19930811  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/925,460  
FILING DATE: August 11, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00246/168001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 214:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
PCT-US93-07545-214

Query Match 1.3%; Score 7; DB 5; Length 26;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
1111111  
Db 13 VTVLINS 19

RESULT 28  
US-08-673-269-7  
Sequence 7, Application US/08673269  
Patent No. 5736344  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OKUBO, Shoichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,269  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728169  
FILING DATE: 10-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440

```

;
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 37 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-673-269-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 1: Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38
Db 1 CSCSSWL 7

RESULT 29
US-09-192-048-19
; Sequence 19, Application US/09192048
; Patent No. 6063592
; GENERAL INFORMATION:
; APPLICANT: Lee, Sohee
; TITLE OF INVENTION: Kell Protein Proteolytic Activity
; FILE REFERENCE: Sequence 1-27
; Patent No. 6063592
; CURRENT APPLICATION NUMBER: US/09/192,048
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 19
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-192-048-19

Query Match
Best Local Similarity 100.0%; Score 7; DB 3: Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38
Db 1 CSCSSWL 7

RESULT 30
US-07-728-169-7
; Sequence 7, Application US/07728169
; Patent No. 6365725
; GENERAL INFORMATION:
; APPLICANT: ONDA, Haruo
; APPLICANT: OHKUBO, Shoichi
; APPLICANT: KOSAKA, Takuo
; TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE
; TITLE OF INVENTION: PEPTIDE AND USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentl Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/728,169
; FILING DATE: 19910710
```

```

;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: RESNIK, David S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 41109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 37 amino acids
;   TYPE: AMINO ACID
;   TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-728-169-7

Query Match
Best Local Similarity 100.0%; Score 7; DB 4: Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38
Db 1 CSCSSWL 7

RESULT 31
US-09-192-048-19
; Sequence 19, Application US/09192048
; Patent No. 6063592
; GENERAL INFORMATION:
; APPLICANT: Lee, Sohee
; TITLE OF INVENTION: Kell Protein Proteolytic Activity
; FILE REFERENCE: Sequence 1-27
; Patent No. 6063592
; CURRENT APPLICATION NUMBER: US/09/192,048
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 19
; LENGTH: 37
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-192-048-19

Query Match
Best Local Similarity 100.0%; Score 7; DB 6: Length 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38
Db 12 CSCSSWL 18

RESULT 32
US-08-673-269-8
; Sequence 8, Application US/08673269
; Patent No. 5756344
; GENERAL INFORMATION:
; APPLICANT: ONDA, Haruo
; APPLICANT: OHKUBO, Shoichi
; APPLICANT: KOSAKA, Takuo
; TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE
; TITLE OF INVENTION: PEPTIDE AND USE THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; ADDRESSEE: CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentl Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/673,269  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728169  
FILING DATE: 10-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-673-269-8

Query Match 1.3%; Score 7; DB 1; Length 38;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
|||||  
Db 1 CSCSSWL 7

RESULT 33  
US-07-728-169-8  
Sequence 8, Application US/07728169  
Patent No. 6365725  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OKUBO, Shoichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
TITLE OF INVENTION: PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/728,169  
FILING DATE: 19910710  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-07-728-169-8

Query Match 1.3%; Score 7; DB 4; Length 38;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
|||||  
Db 1 CSCSSWL 7

RESULT 34  
US-08-673-269-2  
Sequence 2, Application US/08673269  
Patent No. 5756344  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo  
APPLICANT: OKUBO, Shoichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
TITLE OF INVENTION: PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESS: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/673,269  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728169  
FILING DATE: 10-JUL-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 178 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-673-269-2

Query Match 1.3%; Score 7; DB 1; Length 178;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
|||||  
Db 49 CSCSSWL 55

RESULT 35  
US-07-728-169-2  
Sequence 2, Application US/07728169  
Patent No. 6365725  
GENERAL INFORMATION:  
APPLICANT: ONDA, Haruo



APPLICANT: OKUBO, Shioichi  
APPLICANT: KOSAKA, Takuo  
TITLE OF INVENTION: DNA CODING FOR A HUMAN VASOCONSTRICTIVE  
TITLE OF INVENTION: PEPTIDE AND USE THEREOF  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &  
ADDRESSEE: CUSHMAN  
STREET: 130 Water Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/728,169  
FILING DATE: 19910710  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNIK, David S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41109  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)523-3400  
TELEFAX: (617)523-6440  
TELEX: 200291 STRE UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 178 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-728-169-2

Query Match 1.3%; Score 7; DB 4; Length 178;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
Db 49 CSCSSWL 55

RESULT 36  
US-08-960-190A-38  
Sequence 38, Application US/08960190A  
Patent No. 6232445  
GENERAL INFORMATION:  
APPLICANT: Rhode, Peter R.  
APPLICANT: Acevedo, Jorge  
APPLICANT: Burkhardt, Martin  
APPLICANT: Jiao, Jin-an  
APPLICANT: Wong, Hing C.  
TITLE OF INVENTION: SOLUBLE MHC COMPLEXES AND  
TITLE OF INVENTION: METHODS OF USE THEREOF  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BronsteIn, Roberts & Cushman, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: usa  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/960,190A  
FILING DATE: 29-OCT-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Corless, Peter F  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 48002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX:  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-960-190A-38

Query Match 1.3%; Score 7; DB 4; Length 192;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLINS 370  
Db 89 VTVLINS 95

RESULT 37  
US-08-470-535-12  
Sequence 12, Application US/08470535  
Patent No. 6090587  
GENERAL INFORMATION:  
APPLICANT: Rhodes, Eric T  
APPLICANT: Nag, Bishwajit  
TITLE OF INVENTION: PROKARYOTIC EXPRESSION OF MHC PROTEINS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Street Tower  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105-1492  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/470,535  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/143,575  
FILING DATE: 25-OCT-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,010  
FILING DATE: 25-OCT-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 14058-21-1-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-470-535-12

Query Match      1.3% Score 7; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      364 VTVLTNS 370
      |||||
Db      89 VTVLTNS 95

RESULT 38
US-09-149-476-485
; Sequence 485, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P202P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
```

```

; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,662
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,872
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,882
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,637
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,903
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,888
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,879
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,880
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,894
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,911
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
```

EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 1.3%; Score 7; DB 4; Length 238;  
Best Local Similarity 100.0%; Pred. No. 84;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 26 LCLLCL 32  
Db 206 LCLLCL 212

RESULT 39  
US-09-248-335-32  
Sequence 32, Application US/09248335  
Patent No. 6096504  
GENERAL INFORMATION:  
APPLICANT: MCGONIGLE, BRIAN  
TITLE OF INVENTION: PLANT GLUTATHIONE-S-TRANSFERASE ENZYMES  
FILE REFERENCE: CL-1128-A  
CURRENT APPLICATION NUMBER: US/09/248,335  
CURRENT FILING DATE: 1999-02-10  
EARLIER APPLICATION NUMBER: 08/924,759  
EARLIER FILING DATE: 1997-September-05  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: Microsoft Word Version 7.0A  
SEQ ID NO 32  
LENGTH: 246  
TYPE: PRT  
ORGANISM: maize  
US-09-248-335-32

Query Match 1.3%; Score 7; DB 3; Length 246;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 442 DPRSARL 448  
Db 204 DPRSARL 210

RESULT 40  
US-08-644-664B-27  
Sequence 27, Application US/08644664B  
Patent No. 5776746  
GENERAL INFORMATION:  
APPLICANT: Denney Jr., Dan W.  
TITLE OF INVENTION: Gene Amplification Methods  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP.  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States Of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/644,664B  
FILING DATE: 01-MAY-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: GENITOP-00912  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 248 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-644-664B-27

Query Match 1.3%; Score 7; DB 1; Length 248;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTVLITNS 370  
|||  
Db 114 VTVLITNS 120

Search completed: May 12, 2003, 10:09:07  
Job time : 35 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:08:52 ; Search time 25 Seconds  
(Without alignments)  
1932.540 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 525  
Sequence: 1 MRANPKQTQAMPSETISLMKT.....KLWKRIAKITSLPIESLL 525

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	1.5	498	9	US-09-738-626-4855
2	7	1.3	29	10	US-09-864-761-40182
3	7	1.3	106	9	US-10-000-2564-146
4	7	1.3	149	9	US-09-738-626-4475
5	7	1.3	173	10	US-09-779-307-8
6	7	1.3	173	10	US-09-779-307-24
7	7	1.3	178	10	US-09-779-307-25
8	7	1.3	178	10	US-09-779-307-26
9	7	1.3	178	10	US-09-872-523-67
10	7	1.3	192	10	US-09-766-378A-38
11	7	1.3	205	9	US-09-764-891-4458
12	7	1.3	229	9	US-10-078-090-168
13	7	1.3	238	9	US-09-808-391-485
14	7	1.3	248	9	US-09-925-664-27
15	7	1.3	272	10	US-09-846-808-4
16	7	1.3	281	10	US-09-925-302-498
17	7	1.3	298	10	US-09-815-242-12434
18	7	1.3	302	10	US-09-925-300-1235
19	7	1.3	320	10	US-09-815-242-5650

20	7	1.3	344	10	US-09-992-647-12	Sequence 12, Appl
21	7	1.3	364	10	US-09-756-983-22	Sequence 22, Appl
22	7	1.3	394	9	US-09-974-298-81	Sequence 81, Appl
23	7	1.3	394	9	US-10-057-832-2	Sequence 2, Appl1
24	7	1.3	413	9	US-09-738-626-4478	Sequence 4478, Ap
25	7	1.3	428	10	US-09-881-752A-356	Sequence 356, App
26	7	1.3	489	9	US-10-245-103-42	Sequence 42, Appl
27	7	1.3	489	9	US-10-245-107-42	Sequence 42, Appl
28	7	1.3	489	9	US-10-245-143-42	Sequence 42, Appl
29	7	1.3	489	9	US-10-245-171-42	Sequence 42, Appl
30	7	1.3	489	9	US-10-245-851-42	Sequence 42, Appl
31	7	1.3	489	9	US-10-245-883-42	Sequence 42, Appl
32	7	1.3	489	9	US-10-237-535-42	Sequence 42, Appl
33	7	1.3	489	9	US-10-238-183-42	Sequence 42, Appl
34	7	1.3	489	9	US-10-238-283-42	Sequence 42, Appl
35	7	1.3	489	9	US-10-238-370-42	Sequence 42, Appl
36	7	1.3	489	9	US-10-245-055-42	Sequence 42, Appl
37	7	1.3	489	9	US-10-245-147-42	Sequence 42, Appl
38	7	1.3	489	9	US-10-245-730-42	Sequence 42, Appl
39	7	1.3	489	9	US-10-245-739-42	Sequence 42, Appl
40	7	1.3	489	9	US-10-246-210-42	Sequence 42, Appl
41	7	1.3	489	9	US-10-239-196-42	Sequence 42, Appl
42	7	1.3	489	9	US-10-243-024-42	Sequence 42, Appl
43	7	1.3	489	9	US-10-243-409-42	Sequence 42, Appl
44	7	1.3	489	9	US-10-245-033-42	Sequence 42, Appl
45	7	1.3	489	9	US-10-245-033-42	Sequence 42, Appl
46	7	1.3	489	9	US-10-245-621-42	Sequence 42, Appl
47	7	1.3	489	9	US-10-245-880-42	Sequence 42, Appl
48	7	1.3	489	9	US-10-243-095-42	Sequence 42, Appl
49	7	1.3	489	9	US-10-245-185-42	Sequence 42, Appl
50	7	1.3	489	9	US-10-245-427-42	Sequence 42, Appl
51	7	1.3	489	9	US-10-245-473-42	Sequence 42, Appl
52	7	1.3	489	9	US-10-245-770-42	Sequence 42, Appl
53	7	1.3	489	9	US-10-245-877-42	Sequence 42, Appl
54	7	1.3	489	9	US-10-246-976-42	Sequence 42, Appl
55	7	1.3	506	12	US-10-010-943-2	Sequence 2, Appl1
56	7	1.3	525	9	US-10-020-038-8	Sequence 8, Appl1
57	7	1.3	525	9	US-09-813-153-269	Sequence 269, App
58	7	1.3	597	9	US-09-996-634-129	Sequence 129, App
59	7	1.3	597	9	US-09-997-181-129	Sequence 129, App
60	7	1.3	597	9	US-09-997-182-129	Sequence 129, App
61	7	1.3	601	10	US-09-871-615-2	Sequence 2, Appl1
62	7	1.3	652	10	US-09-992-647-1	Sequence 1, Appl1
63	7	1.3	661	10	US-10-073-865-114	Sequence 114, App
64	7	1.3	661	10	US-09-764-853-679	Sequence 679, App
65	7	1.3	695	10	US-09-764-998-224	Sequence 224, App
66	7	1.3	1018	9	US-09-738-626-5527	Sequence 5527, App
67	7	1.3	1317	10	US-10-101-464A-909	Sequence 909, App
68	7	1.3	3069	9	US-09-815-242-5118	Sequence 5118, Ap
69	7	1.1	8	10	US-09-712-363-246	Sequence 246, App
70	7	1.1	13	9	US-09-756-283A-51	Sequence 51, Appl
71	7	1.1	18	9	US-09-953-354-130	Sequence 130, App
72	7	1.1	18	9	US-10-084-813-45	Sequence 45, Appl
73	7	1.1	18	9	US-10-084-813-46	Sequence 46, Appl
74	7	1.1	18	9	US-10-084-813-47	Sequence 47, Appl
75	7	1.1	28	10	US-09-864-761-40325	Sequence 40325, A
76	7	1.1	28	10	US-09-864-761-40537	Sequence 40537, A
77	7	1.1	31	9	US-09-933-367A-29	Sequence 29, Appl
78	7	1.1	31	9	US-09-933-367A-33	Sequence 33, Appl
79	7	1.1	31	10	US-09-864-761-36002	Sequence 36002, A
80	7	1.1	40	12	US-10-044-592-80	Sequence 80, Appl
81	7	1.1	45	9	US-09-986-480-336	Sequence 336, App
82	7	1.1	45	9	US-10-091-572-363	Sequence 363, App
83	7	1.1	45	9	US-09-764-891-3306	Sequence 3306, App
84	7	1.1	45	9	US-09-764-891-3808	Sequence 4808, Ap
85	7	1.1	46	10	US-09-897-877-356	Sequence 356, App
86	7	1.1	46	10	US-09-864-816-111	Sequence 111, App
87	7	1.1	47	9	US-09-808-391-523	Sequence 391, App
88	7	1.1	48	10	US-09-864-761-36693	Sequence 36693, A
89	7	1.1	49	10	US-09-864-761-33868	Sequence 33868, A
90	7	1.1	49	10	US-09-864-761-37093	Sequence 37093, A
91	7	1.1	49	10	US-09-733-183A-7	Sequence 7, Appl1
92	7	1.1	49	10	US-09-733-183A-19	Sequence 19, Appl
93	7	1.1	49	10	US-09-733-183A-31	Sequence 31, Appl















969	6	1.1	334	9	US-10-180-547-176	Sequence 176, App
970	6	1.1	334	9	US-10-180-549-176	Sequence 176, App
971	6	1.1	334	9	US-10-180-555-176	Sequence 176, App
972	6	1.1	334	9	US-10-180-559-176	Sequence 176, App
973	6	1.1	334	9	US-10-181-000-176	Sequence 176, App
974	6	1.1	334	9	US-10-183-010-176	Sequence 176, App
975	6	1.1	334	9	US-10-183-012-176	Sequence 176, App
976	6	1.1	334	9	US-10-184-614-176	Sequence 176, App
977	6	1.1	334	9	US-10-184-623-176	Sequence 176, App
978	6	1.1	334	9	US-10-184-635-176	Sequence 176, App
979	6	1.1	334	9	US-10-184-637-176	Sequence 176, App
980	6	1.1	334	9	US-10-184-646-176	Sequence 176, App
981	6	1.1	334	9	US-10-184-647-176	Sequence 176, App
982	6	1.1	334	9	US-10-184-652-176	Sequence 176, App
983	6	1.1	334	9	US-10-187-594-176	Sequence 176, App
984	6	1.1	334	9	US-10-187-596-176	Sequence 176, App
985	6	1.1	334	9	US-10-187-745-176	Sequence 176, App
986	6	1.1	334	9	US-10-187-885-176	Sequence 176, App
987	6	1.1	334	9	US-10-187-886-176	Sequence 176, App
988	6	1.1	334	9	US-10-199-464-176	Sequence 176, App
989	6	1.1	334	9	US-10-176-751-176	Sequence 176, App
990	6	1.1	334	9	US-10-176-760-176	Sequence 176, App
991	6	1.1	334	9	US-10-176-990-176	Sequence 176, App
992	6	1.1	334	9	US-10-180-541-176	Sequence 176, App
993	6	1.1	334	9	US-10-180-542-176	Sequence 176, App
994	6	1.1	334	9	US-10-180-548-176	Sequence 176, App
995	6	1.1	334	9	US-10-180-551-176	Sequence 176, App
996	6	1.1	334	9	US-10-180-998-176	Sequence 176, App
997	6	1.1	334	9	US-10-180-999-176	Sequence 176, App
998	6	1.1	334	9	US-10-183-013-176	Sequence 176, App
999	6	1.1	334	9	US-10-184-612-176	Sequence 176, App
1000	6	1.1	334	9	US-10-184-616-176	Sequence 176, App

## ALIGNMENTS

RESULT 1  
US-09-738-626-4855  
Sequence 4855, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SEMOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 4855  
LENGTH: 498  
TYPE: PRT  
ORGANISM: *Corynebacterium glutamicum*  
US-09-738-626-4855

Query Match 1.5%; Score 8; DB 9; Length 498;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 440 NLDPRSAR 447  
Db 418 NLDPRSAR 425

RESULT 2  
US-09-864-761-40182  
Sequence 40182, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Acomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864, 761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180, 312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207, 456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632, 366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263, 6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236, 359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234, 687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608, 408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774, 203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 40182  
LENGTH: 29  
TYPE: PRT  
ORGANISM: *Homo sapiens*  
FEATURE:  
OTHER INFORMATION: MAP TO AL157381.7  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93  
US-09-864-761-40182

Query Match 1.3%; Score 7; DB 10; Length 29;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 455 VIESPKI 461  
|||||||  
DB 7 VIESPKI 13

## RESULT 3

US-10-000-256A-146  
; Sequence 146, Application US/10000256A  
; Publication No. US20030039983A1  
; GENERAL INFORMATION:  
; APPLICANT: Sun, Yongming  
; APPLICANT: Recipon, Heirve  
; APPLICANT: Chen, Sei-yu  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and  
; FILE REFERENCE: DEX-0259  
; CURRENT APPLICATION NUMBER: US/10/000,256A  
; PRIOR FILING DATE: 2001-11-01  
; PRIOR APPLICATION NUMBER: 60/244,782  
; PRIOR FILING DATE: 2000-11-01  
; NUMBER OF SEQ ID NOS: 240  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 146  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-10-000-256A-146

Query Match 1.3%; Score 7; DB 9; Length 106;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 416 GSSVTSI 422  
|||||||  
DB 43 GSSVTSI 49

## RESULT 4

US-09-738-626-4475  
; Sequence 4475, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4475  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4475

Query Match 1.3%; Score 7; DB 9; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SLICLLI 31  
|||||||  
DB 14 SLICLLI 20

## RESULT 5

US-09-779-307-8  
; Sequence 8, Application US/09779307  
; Patent No. US20020137675A1  
; GENERAL INFORMATION:  
; APPLICANT: Taupier Jr., Raymond  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Verneet, Corine  
; APPLICANT: Prayaga, Sudhirdas  
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-662 US  
; CURRENT APPLICATION NUMBER: US/09/779,307  
; PRIOR FILING DATE: 2001-02-07  
; PRIOR APPLICATION NUMBER: 60/180,880  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,044  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,656  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: 60/182,795  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-779-307-8

Query Match 1.3%; Score 7; DB 10; Length 173;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
|||||||  
DB 49 CSCSSWL 55

## RESULT 6

US-09-779-307-24  
; Sequence 24, Application US/09779307  
; Patent No. US20020137675A1  
; GENERAL INFORMATION:  
; APPLICANT: Taupier Jr., Raymond  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Verneet, Corine  
; APPLICANT: Prayaga, Sudhirdas  
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby  
; FILE REFERENCE: 15966-662 US  
; CURRENT APPLICATION NUMBER: US/09/779,307  
; PRIOR FILING DATE: 2001-02-07  
; PRIOR APPLICATION NUMBER: 60/180,880  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,044  
; PRIOR FILING DATE: 2000-02-08  
; PRIOR APPLICATION NUMBER: 60/181,656  
; PRIOR FILING DATE: 2000-02-10  
; PRIOR APPLICATION NUMBER: 60/182,795  
; PRIOR FILING DATE: 2000-02-15  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 173

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-779-307-24

```

```

Query Match
Best Local Similarity 1.3%; Score 7; DB 10; Length 173;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 32 CSCSSWL 38
Db 49 CSCSSWL 55

```

```

RESULT 7
US-09-779-307-25

```

```

; Sequence 25, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Prayaga, Sudhirdas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-662 US
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: US/09/779,307
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-10
; PRIOR FILING DATE: 2000-02-10
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-779-307-25

```

```

Query Match
Best Local Similarity 1.3%; Score 7; DB 10; Length 178;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 32 CSCSSWL 38
Db 49 CSCSSWL 55

```

```

RESULT 8
US-09-779-307-26
; Sequence 26, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Vernet, Corine
; APPLICANT: Prayaga, Sudhirdas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-662 US
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: US/09/779,307
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-08
; PRIOR FILING DATE: 2000-02-10
; PRIOR FILING DATE: 2000-02-10
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 43

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-779-307-26

```

```

Query Match
Best Local Similarity 1.3%; Score 7; DB 10; Length 178;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 32 CSCSSWL 38
Db 49 CSCSSWL 55

```

```

RESULT 9
US-09-872-523-67

```

```

; Sequence 67, Application US/09872523
; Patent No. US20020137906A1
; GENERAL INFORMATION:
; APPLICANT: Horvitz, H. Robert
; APPLICANT: Davison, Ewa M.
; APPLICANT: Lu, Xiaowei
; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans
; FILE REFERENCE: 01997/536002
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/208,802
; PRIOR FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-872-523-67

```

```

Query Match
Best Local Similarity 1.3%; Score 7; DB 10; Length 178;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 166 RKWRALG 172
Db 75 RKWRALG 81

```

```

RESULT 10
US-09-766-378A-38
; Sequence 38, Application US/09766378A
; Patent No. US2002091079A1
; GENERAL INFORMATION:
; APPLICANT: Rhode, Peter R.
; APPLICANT: Acevedo, Jorge
; APPLICANT: Burkhardt, Martin
; APPLICANT: Jiao, Jin-an
; APPLICANT: Wong, Ning C.
; TITLE OF INVENTION: SOLUBLE MHC COMPLEXES AND
; METHODS OF USE THEREOF
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESS: Bronstein, Roberts & Cushman, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: usa
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/766,378A  
FILING DATE: 19-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/960,190  
FILING DATE: 29-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Corless, Peter F  
REGISTRATION NUMBER: 33,860  
REFERENCE/DOCKET NUMBER: 48002-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 192 amino acids  
TYPE: amino acid  
STRADEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 38:  
US-09-766-378A-38

Query Match 1.3%: Score 7; DB 10; Length 192;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTULTNS 370  
|||||  
DB 89 VTULTNS 95

RESULT 11  
US-09-764-891-4458  
Sequence 4458, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 4458  
LENGTH: 205  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (153)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (190)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-891-4458

Query Match 1.3%: Score 7; DB 9; Length 205;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 KPLKAG 394  
|||||  
DB 193 KPLKAG 199

RESULT 12  
US-10-078-090-168  
Sequence 168, Application US/10078090  
Publication No. US20030044815A1  
GENERAL INFORMATION:

APPLICANT: Salceda, Susana  
APPLICANT: Macina, Roberto  
APPLICANT: Hu, Ping  
APPLICANT: Recipon, Herve  
APPLICANT: Karra, Kalpana  
APPLICANT: Caffrey, Robert  
APPLICANT: Sun, Yongming  
APPLICANT: Liu, Chenghua  
TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and  
FILE REFERENCE: DEX-0312  
CURRENT APPLICATION NUMBER: US/10/078,090  
CURRENT FILING DATE: 2002-02-14  
PRIOR APPLICATION NUMBER: 60/268,999  
PRIOR FILING DATE: 2001-02-15  
NUMBER OF SEQ ID NOS: 210  
SOFTWARE: Patentln version 3.1  
SEQ ID NO 168  
LENGTH: 229  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-078-090-168

Query Match 1.3%: Score 7; DB 9; Length 229;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SILCLL 31  
|||||  
DB 2 SILCLL 8

RESULT 13  
US-09-809-391-485  
Sequence 485, Application US/09809391  
Publication No. US20030049618A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: PZ002P2  
CURRENT APPLICATION NUMBER: US/09/809,391  
CURRENT FILING DATE: 2001-03-16  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 761  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 485  
LENGTH: 238  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (11)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (14)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (22)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (63)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (64)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (66)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-809-391-485

Query Match 1.3%: Score 7; DB 9; Length 238;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 LCLC 32  
|||||||  
DB 206 LCLC 212

RESULT 14  
US-09-925-664-27  
; Sequence 27, Application US/09925664  
; Patent No. US2002016006A1  
; GENERAL INFORMATION:  
; APPLICANT: Denney, Jr., Dan W.  
; TITLE OF INVENTION: Methods of Treating Lymphoma and Leukemia  
; FILE REFERENCE: GENOTYPE-06499  
; CURRENT APPLICATION NUMBER: US/09/925,664  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 09/370,453  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: 08/644,664  
; PRIOR FILING DATE: 1996-05-01  
; PRIOR APPLICATION NUMBER: 08/761,277  
; PRIOR FILING DATE: 1996-12-06  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-925-664-27

Query Match  
Best Local Similarity 1.3%; Score 7; DB 9; Length 248;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTULTNS 370  
|||||||  
DB 114 VTULTNS 120

RESULT 15  
US-09-846-808-4  
; Sequence 4, Application US/09846808  
; Patent No. US20020064531A1  
; GENERAL INFORMATION:  
; APPLICANT: Walker, David H.  
; APPLICANT: Yu, Xu-Jie  
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane  
; FILE REFERENCE: D6311  
; CURRENT APPLICATION NUMBER: US/09/846,808  
; PRIOR FILING DATE: 2001-05-01  
; PRIOR APPLICATION NUMBER: 60/201,035  
; PRIOR FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 53  
; SEQ ID NO 4  
; LENGTH: 272  
; TYPE: PRT  
; ORGANISM: Ehrlichia chaffeensis  
; FEATURE:  
; OTHER INFORMATION: P28-4 Outer Membrane Protein of  
; OTHER INFORMATION: Ehrlichia chaffeensis  
US-09-846-808-4

Query Match  
Best Local Similarity 1.3%; Score 7; DB 10; Length 272;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 LPIESL 524  
|||||||  
DB 23 LPIESL 29

RESULT 16  
US-09-925-302-498  
; Sequence 498, Application US/09925302  
; Patent No. US20020044941A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA104  
; CURRENT APPLICATION NUMBER: US/09/925,302  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05918  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 896  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 498  
; LENGTH: 281  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (5)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (21)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-498

Query Match  
Best Local Similarity 1.3%; Score 7; DB 10; Length 281;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VTULTNS 370  
|||||||  
DB 141 VTULTNS 147

RESULT 17  
US-09-815-242-12434  
; Sequence 12434, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA-011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12434



LENGTH: 298  
 TYPE: PRT  
 ORGANISM: Staphylococcus aureus  
 US-09-815-242-12434

Query Match 1.3%; Score 7; DB 10; Length 298;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 GYNDETS 270  
 DB 127 GYNDETS 133

RESULT 18  
 US-09-925-300-1235  
 Sequence 1235, Application US/09925300  
 Patent No. US20020151681A1  
 GENERAL INFORMATION:  
 APPLICANT: Craig Rosen,  
 APPLICANT: Steve Ruben,  
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 FILE REFERENCE: PA101  
 CURRENT APPLICATION NUMBER: US/09/925,300  
 PRIOR FILING DATE: 2001-08-10  
 PRIOR APPLICATION NUMBER: PCT/US00/05988  
 PRIOR FILING DATE: 2000-03-08  
 PRIOR APPLICATION NUMBER: 60/124,270  
 PRIOR FILING DATE: 1998-03-12  
 NUMBER OF SEQ ID NOS: 1890  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1235  
 LENGTH: 302  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (2)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (226)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-925-300-1235

Query Match 1.3%; Score 7; DB 10; Length 302;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 511 IAAKILS 517  
 DB 149 IAAKILS 155

RESULT 19  
 US-09-815-242-5650  
 Sequence 5650, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 5650  
 LENGTH: 320  
 TYPE: PRT  
 ORGANISM: Staphylococcus aureus  
 US-09-815-242-5650

Query Match 1.3%; Score 7; DB 10; Length 320;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 GYNDETS 270  
 DB 149 GYNDETS 155

RESULT 20  
 US-09-992-647-12  
 Sequence 12, Application US/09992647  
 Patent No. US20020146767A1  
 GENERAL INFORMATION:  
 APPLICANT: Xu, Hong  
 APPLICANT: Cohan, Victoria L.  
 APPLICANT: Stuan, Susan G.  
 TITLE OF INVENTION: HUMAN EMR1-LIKE G PROTEIN-COUPLED RECEPTOR  
 FILE REFERENCE: PC-0052 CIP  
 CURRENT APPLICATION NUMBER: US/09/992,647  
 CURRENT FILING DATE: 2002-11-13  
 NUMBER OF SEQ ID NOS: 12  
 SOFTWARE: PERL Program  
 SEQ ID NO 12  
 LENGTH: 344  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 PUBLICATION INFORMATION:  
 DATABASE ACCESSION NUMBER: Genbank ID No. US20020146767A1 92935597  
 US-09-992-647-12

Query Match 1.3%; Score 7; DB 10; Length 344;  
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SLICLL 31  
 DB 68 SLICLL 74

RESULT 21  
 US-09-756-983-22  
 Sequence 22, Application US/09756983  
 Patent No. US20020122818A1  
 GENERAL INFORMATION:  
 APPLICANT: Albani, Salvatore  
 TITLE OF INVENTION: METHODS FOR ISOLATION, QUANTIFICATION,  
 CHARACTERIZATION AND MODULATION OF  
 TITLE OF INVENTION: ANTIGEN-SPECIFIC T CELLS  
 FILE REFERENCE: 246/285-CIP  
 CURRENT APPLICATION NUMBER: US/09/756,983  
 CURRENT FILING DATE: 2001-01-09  
 PRIOR APPLICATION NUMBER: 60/105,018

```

; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 09/421,506
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: PCT/US99/2466
; PRIOR FILING DATE: 1999-10-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 22
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Fusion constructs with human and bacterial sequences
US-09-756-983-22
```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 364;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 364 VIVLINS 370
DB 114 VIVLINS 120
```

```

RESULT 22
US-09-974-298-81
; Sequence 81, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huel-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO: 81
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 1969731CD1
US-09-974-298-81
```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 394;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 416 GSSVTSLS 422
DB 331 GSSVTSLS 337
```

```

RESULT 23
US-10-057-832-2
; Sequence 2, Application US/10057832
; Publication No. US20020182621A1
; GENERAL INFORMATION:
; APPLICANT: Costa, Max
; APPLICANT: Salnikow, Konstantin
; APPLICANT: Yee, Herman
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING CAP43 PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 5986/11147US1
; CURRENT APPLICATION NUMBER: US/10/057,832
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 10/057,832
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
```

```

; SEQ ID NO: 2
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-057-832-2
```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 394;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 416 GSSVTSLS 422
DB 331 GSSVTSLS 337
```

```

RESULT 24
US-09-738-626-4478
; Sequence 4478, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO: 4478
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4478
```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 413;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 223 ATGSVVG 229
DB 76 ATGSVVG 82
```

```

RESULT 25
US-09-881-752A-356
; Sequence 356, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in t
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
```

```
;; PRIOR APPLICATION NUMBER: US 08/833,457
;; PRIOR FILING DATE: 1997-04-01
;; NUMBER OF SEQ ID NOS: 370
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 356
;; LENGTH: 428
;; TYPE: PRT
;; ORGANISM: Helicobacter pylori
US-09-881-752A-356
```

```
Query Match          1.3%; Score 7; DB 10; Length 428;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 199 GGRNIGD 205
    |||||
DB 107 GGRNIGD 113
```

RESULT 26

```
;; Sequence 42, Application US/10245103
;; Publication No. US20030068778a1
;; GENERAL INFORMATION:
```

```
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watande, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C12
;; CURRENT APPLICATION NUMBER: US/10/245,103
;; CURRENT FILING DATE: 2002-09-17
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-103-42
```

```
Query Match          1.3%; Score 7; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 127 GVRVRL 133
    |||||
DB 353 GVRVRL 359
```

RESULT 27

```
;; Sequence 42, Application US/10245107
;; Publication No. US20030068779a1
;; GENERAL INFORMATION:
```

```
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Grimaldi, J. Christopher
;; APPLICANT: Gurney, Austin
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stephan, Jean-Phillippe
;; APPLICANT: Watande, Colin
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; APPLICANT: Fong, Sherman
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3630R1C71
;; CURRENT APPLICATION NUMBER: US/10/245,107
;; CURRENT FILING DATE: 2002-09-16
;; PRIOR APPLICATION NUMBER: 10/197942
;; PRIOR FILING DATE: 2002-07-18
;; PRIOR APPLICATION NUMBER: 60/059114
;; PRIOR FILING DATE: 1997-09-17
;; PRIOR APPLICATION NUMBER: 60/063046
;; PRIOR FILING DATE: 1997-10-24
;; PRIOR APPLICATION NUMBER: 60/065027
;; PRIOR FILING DATE: 1997-11-10
;; PRIOR APPLICATION NUMBER: 60/079689
;; PRIOR FILING DATE: 1998-03-27
;; PRIOR APPLICATION NUMBER: 60/086478
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090689
;; PRIOR FILING DATE: 1998-06-25
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 116
;; SEQ ID NO 42
;; LENGTH: 489
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-245-107-42
```

```
Query Match          1.3%; Score 7; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 127 GVRVRL 133
    |||||
DB 353 GVRVRL 359
```

RESULT 28

```
;; Sequence 42, Application US/10245143
;; Publication No. US20030068780a1
;; GENERAL INFORMATION:
```

```
;; APPLICANT: Baker, Kevin
;; APPLICANT: Eaton, Dan
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Goddard, Audrey
```

```

: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Guiney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Philippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630R1C90
: CURRENT APPLICATION NUMBER: US/10/245,143
: PRIOR FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 42
: LENGTH: 489
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-245-143-42

Query Match      1.3%; Score 7; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      127 GVRVRL 133
Db      353 GVRVRL 359

RESULT 29
US-10-245-771-42
: Sequence 42, Application US/10245771
: Publication No. US20030068781A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Baton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Guiney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Philippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630R1C98
: CURRENT APPLICATION NUMBER: US/10/245,771
: PRIOR FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: 10/197942
```

```

: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: PRIOR FILING DATE: 1998-06-25
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 42
: LENGTH: 489
: TYPE: PRT
: ORGANISM: Homo Sapien
: US-10-245-771-42

Query Match      1.3%; Score 7; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      127 GVRVRL 133
Db      353 GVRVRL 359

RESULT 30
US-10-245-851-42
: Sequence 42, Application US/10245851
: Publication No. US20030068782A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Baton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Guiney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Philippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630R1C93
: CURRENT APPLICATION NUMBER: US/10/245,851
: PRIOR FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: 10/197942
: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
```

;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 116  
;; SEQ ID NO 42  
;; LENGTH: 489  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-245-851-42

Query Match  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 GVRRL 133  
Db 353 GVRRL 359

## RESULT 31

US-10-245-883-42  
;; Sequence 42, Application US/10245883  
;; Publication No. US20030068783A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Baker, Kevin  
;; APPLICANT: Fliviaroff, Ellen  
;; APPLICANT: Goddard, Audrey  
;; APPLICANT: Grimaldi, J. Christopher  
;; APPLICANT: Gurney, Austin  
;; APPLICANT: Smith, Victoria  
;; APPLICANT: Stephan, Jean-Phillippe  
;; APPLICANT: Watanabe, Colin  
;; APPLICANT: Wood, William  
;; APPLICANT: Zhang, Zemin  
;; APPLICANT: Fong, Sherman  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; FILE REFERENCE: P3630R1C70  
;; CURRENT APPLICATION NUMBER: US/10/245, 883  
;; CURRENT FILING DATE: 2002-09-16  
;; PRIOR APPLICATION NUMBER: 10/197942  
;; PRIOR FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/059114  
;; PRIOR FILING DATE: 1997-09-17  
;; PRIOR APPLICATION NUMBER: 60/063046  
;; PRIOR FILING DATE: 1997-10-24  
;; PRIOR APPLICATION NUMBER: 60/065027  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087607  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 116  
;; SEQ ID NO 42  
;; LENGTH: 489  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-245-883-42

Query Match 1.3%; Score 7; DB 9; Length 489;

Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 GVRRL 133  
Db 353 GVRRL 359

## RESULT 32

US-10-237-535-42  
;; Sequence 42, Application US/10237535  
;; Publication No. US20030073188A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Baker, Kevin  
;; APPLICANT: Fliviaroff, Ellen  
;; APPLICANT: Goddard, Audrey  
;; APPLICANT: Grimaldi, J. Christopher  
;; APPLICANT: Gurney, Austin  
;; APPLICANT: Smith, Victoria  
;; APPLICANT: Stephan, Jean-Phillippe  
;; APPLICANT: Watanabe, Colin  
;; APPLICANT: Wood, William  
;; APPLICANT: Zhang, Zemin  
;; APPLICANT: Fong, Sherman  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; FILE REFERENCE: P3630R1C3  
;; CURRENT APPLICATION NUMBER: US/10/237, 535  
;; CURRENT FILING DATE: 2002-09-06  
;; PRIOR APPLICATION NUMBER: 10/197942  
;; PRIOR FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/059114  
;; PRIOR FILING DATE: 1997-09-17  
;; PRIOR APPLICATION NUMBER: 60/063046  
;; PRIOR FILING DATE: 1997-10-24  
;; PRIOR APPLICATION NUMBER: 60/065027  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087607  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/089801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; PRIOR APPLICATION NUMBER: 60/091358  
;; PRIOR FILING DATE: 1998-07-01  
;; PRIOR APPLICATION NUMBER: 60/091978  
;; PRIOR FILING DATE: 1998-07-07  
;; PRIOR APPLICATION NUMBER: 60/099803  
;; PRIOR FILING DATE: 1998-09-10  
;; PRIOR APPLICATION NUMBER: 60/106932  
;; PRIOR FILING DATE: 1998-11-03  
;; PRIOR APPLICATION NUMBER: 60/115554  
;; PRIOR FILING DATE: 1999-01-12  
;; PRIOR APPLICATION NUMBER: 60/119342  
;; PRIOR FILING DATE: 1999-02-09  
;; PRIOR APPLICATION NUMBER: 60/129957  
;; PRIOR FILING DATE: 1999-03-12  
;; PRIOR APPLICATION NUMBER: 60/123972  
;; PRIOR FILING DATE: 1999-03-11  
;; PRIOR APPLICATION NUMBER: 60/127372  
;; PRIOR FILING DATE: 1999-04-01  
;; PRIOR APPLICATION NUMBER: 60/131271  
;; PRIOR FILING DATE: 1999-04-27  
;; PRIOR APPLICATION NUMBER: 60/133459  
;; PRIOR FILING DATE: 1999-05-11  
;; PRIOR APPLICATION NUMBER: 60/135725

;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/135729  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/135750  
;; PRIOR FILING DATE: 1999-05-25  
;; PRIOR APPLICATION NUMBER: 60/138385  
;; PRIOR FILING DATE: 1999-06-09  
;; PRIOR APPLICATION NUMBER: 60/140653  
;; PRIOR FILING DATE: 1999-06-22  
;; PRIOR APPLICATION NUMBER: 60/141037  
;; PRIOR FILING DATE: 1999-06-23  
;; PRIOR APPLICATION NUMBER: 60/144732  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/144758  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/144790  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/145228  
;; PRIOR FILING DATE: 1999-07-20  
;; PRIOR APPLICATION NUMBER: 60/145698  
;; PRIOR FILING DATE: 1999-07-26  
;; PRIOR APPLICATION NUMBER: 60/146222  
;; PRIOR FILING DATE: 1999-07-28  
;; PRIOR APPLICATION NUMBER: 60/146843  
;; PRIOR FILING DATE: 1999-08-03  
;; PRIOR APPLICATION NUMBER: 60/148188  
;; PRIOR FILING DATE: 1999-08-10  
;; PRIOR APPLICATION NUMBER: 60/148513  
;; PRIOR FILING DATE: 1999-08-12  
;; PRIOR APPLICATION NUMBER: 60/149327  
;; PRIOR FILING DATE: 1999-08-17  
;; PRIOR APPLICATION NUMBER: 60/149395  
;; PRIOR FILING DATE: 1999-08-17  
;; PRIOR APPLICATION NUMBER: 60/150114  
;; PRIOR FILING DATE: 1999-08-20  
;; PRIOR APPLICATION NUMBER: 60/151700  
;; PRIOR FILING DATE: 1999-08-31  
;; PRIOR APPLICATION NUMBER: 60/151734  
;; PRIOR FILING DATE: 1999-08-31  
;; PRIOR APPLICATION NUMBER: 60/162506  
;; PRIOR FILING DATE: 1999-10-29  
;; PRIOR APPLICATION NUMBER: 60/170262  
;; PRIOR FILING DATE: 1999-12-09  
;; PRIOR APPLICATION NUMBER: 60/177118  
;; PRIOR FILING DATE: 2000-01-20  
;; PRIOR APPLICATION NUMBER: 60/179851  
;; PRIOR FILING DATE: 2000-02-02  
;; PRIOR APPLICATION NUMBER: 60/180921  
;; PRIOR FILING DATE: 2000-02-08  
;; PRIOR APPLICATION NUMBER: 60/187202  
;; PRIOR FILING DATE: 2000-03-03  
;; PRIOR APPLICATION NUMBER: 60/198587  
;; PRIOR FILING DATE: 2000-04-18  
;; PRIOR APPLICATION NUMBER: 60/199614  
;; PRIOR FILING DATE: 2000-04-25  
;; PRIOR APPLICATION NUMBER: 60/206330  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/206368  
;; PRIOR FILING DATE: 2000-05-23  
;; PRIOR APPLICATION NUMBER: 60/209832  
;; PRIOR FILING DATE: 2000-06-05  
;; PRIOR APPLICATION NUMBER: 60/218371  
;; PRIOR FILING DATE: 2000-07-13  
;; PRIOR APPLICATION NUMBER: 60/222695  
;; PRIOR FILING DATE: 2000-08-02  
;; PRIOR APPLICATION NUMBER: 60/229896  
;; PRIOR FILING DATE: 2000-09-01  
;; PRIOR APPLICATION NUMBER: 60/230621  
;; PRIOR FILING DATE: 2000-09-05  
;; PRIOR APPLICATION NUMBER: 60/232887  
;; PRIOR FILING DATE: 2000-09-15  
;; PRIOR APPLICATION NUMBER: 60/235147  
;; PRIOR FILING DATE: 2000-09-22

;; PRIOR APPLICATION NUMBER: 60/261878  
;; PRIOR FILING DATE: 2001-01-12  
;; PRIOR APPLICATION NUMBER: 60/261910  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/261939  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/262150  
;; PRIOR FILING DATE: 2001-01-16  
;; PRIOR APPLICATION NUMBER: 60/264395  
;; PRIOR FILING DATE: 2001-01-25  
;; PRIOR APPLICATION NUMBER: 60/266421  
;; PRIOR FILING DATE: 2001-02-02  
;; PRIOR APPLICATION NUMBER: 60/267623  
;; PRIOR FILING DATE: 2001-02-09  
;; PRIOR APPLICATION NUMBER: 60/274399  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 60/280982  
;; PRIOR FILING DATE: 2001-04-03  
;; PRIOR APPLICATION NUMBER: 60/282129  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/282199  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/290589  
;; PRIOR FILING DATE: 2001-05-09  
;; PRIOR APPLICATION NUMBER: 09/180997  
;; PRIOR FILING DATE: 1998-11-19  
;; PRIOR APPLICATION NUMBER: 09/267213  
;; PRIOR FILING DATE: 1999-03-12  
;; PRIOR APPLICATION NUMBER: 09/380137  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 09/380138  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: 09/403297  
;; PRIOR FILING DATE: 1999-10-18  
;; PRIOR APPLICATION NUMBER: 09/423741  
;; PRIOR FILING DATE: 1999-11-10  
;; PRIOR APPLICATION NUMBER: 09/709238  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 09/802706  
;; PRIOR FILING DATE: 2001-03-09  
;; PRIOR APPLICATION NUMBER: 09/872035  
;; PRIOR FILING DATE: 2001-06-01  
;; PRIOR APPLICATION NUMBER: 09/918585  
;; PRIOR FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: 09/924419  
;; PRIOR FILING DATE: 2001-08-06  
;; PRIOR APPLICATION NUMBER: 09/927796  
;; PRIOR FILING DATE: 2001-08-09  
;; PRIOR APPLICATION NUMBER: 09/929404  
;; PRIOR FILING DATE: 2001-08-13  
;; PRIOR APPLICATION NUMBER: 09/931836  
;; PRIOR FILING DATE: 2001-08-16  
;; PRIOR APPLICATION NUMBER: 09/941992  
;; PRIOR FILING DATE: 2001-08-28  
;; PRIOR APPLICATION NUMBER: 09/946374  
;; PRIOR FILING DATE: 2001-09-04  
;; PRIOR APPLICATION NUMBER: 10/001054  
;; PRIOR FILING DATE: 2001-11-30  
;; PRIOR APPLICATION NUMBER: 10/052586  
;; PRIOR FILING DATE: 2002-01-15  
;; PRIOR APPLICATION NUMBER: 10/081056  
;; PRIOR FILING DATE: 2002-02-20  
;; PRIOR APPLICATION NUMBER: 10/119480  
;; PRIOR FILING DATE: 2002-04-09

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 GVRVRL 133  
|||||  
Db 353 GVRVRL 359

RESULT 33  
US-10-238-183-42  
Sequence 42, Application US/10238183  
Publication No. US20030073189A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin  
APPLICANT: Eaton, Dan  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Goddard, Audrey  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gutney, Austin  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Phillippe  
APPLICANT: Matande, Colin  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
APPLICANT: Fong, Sherman  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3630R1C11  
CURRENT APPLICATION NUMBER: US/10/238,183  
CURRENT FILING DATE: 2002-09-09  
PRIOR APPLICATION NUMBER: 10/197942  
PRIOR FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/059114  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/065027  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/086478  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090689  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/091358  
PRIOR FILING DATE: 1998-07-01  
PRIOR APPLICATION NUMBER: 60/091978  
PRIOR FILING DATE: 1998-07-07  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/106932  
PRIOR FILING DATE: 1998-11-03  
PRIOR APPLICATION NUMBER: 60/115554  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/119342  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: 60/123957  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: 60/123972  
PRIOR FILING DATE: 1999-03-11  
PRIOR APPLICATION NUMBER: 60/127372  
PRIOR FILING DATE: 1999-04-01  
PRIOR APPLICATION NUMBER: 60/131271  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/133459  
PRIOR FILING DATE: 1999-05-11  
PRIOR APPLICATION NUMBER: 60/135725  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/135729  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/135750  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/138385  
PRIOR FILING DATE: 1999-06-09

PRIOR APPLICATION NUMBER: 60/140653  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144732  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/144790  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/145228  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/145698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: 60/146222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: 60/146843  
PRIOR FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: 60/148188  
PRIOR FILING DATE: 1999-08-10  
PRIOR APPLICATION NUMBER: 60/148513  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/149327  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/149395  
PRIOR FILING DATE: 1999-08-17  
PRIOR APPLICATION NUMBER: 60/150114  
PRIOR FILING DATE: 1999-08-20  
PRIOR APPLICATION NUMBER: 60/151700  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: 60/151734  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170262  
PRIOR FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: 60/177118  
PRIOR FILING DATE: 2000-01-20  
PRIOR APPLICATION NUMBER: 60/179851  
PRIOR FILING DATE: 2000-02-02  
PRIOR APPLICATION NUMBER: 60/180921  
PRIOR FILING DATE: 2000-02-08  
PRIOR APPLICATION NUMBER: 60/187202  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: 60/198587  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: 60/199614  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 60/206330  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/206368  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/209832  
PRIOR FILING DATE: 2000-06-05  
PRIOR APPLICATION NUMBER: 60/218371  
PRIOR FILING DATE: 2000-07-13  
PRIOR APPLICATION NUMBER: 60/222695  
PRIOR FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: 60/229896  
PRIOR FILING DATE: 2000-09-01  
PRIOR APPLICATION NUMBER: 60/230621  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/233887  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 60/235147  
PRIOR FILING DATE: 2000-09-22  
PRIOR APPLICATION NUMBER: 60/261878  
PRIOR FILING DATE: 2001-01-12  
PRIOR APPLICATION NUMBER: 60/261910  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: 60/261939  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: 60/262150

```

; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/264395
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 60/266421
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/267623
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/274399
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/280982
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282129
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/282199
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/290589
; PRIOR FILING DATE: 2001-05-09
; PRIOR APPLICATION NUMBER: 09/180997
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: 09/267213
; PRIOR FILING DATE: 1999-03-12
; PRIOR APPLICATION NUMBER: 09/380137
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/380138
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/403297
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: 09/423741
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: 09/709238
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 09/802706
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 09/872035
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 09/918585
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 09/924419
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/927796
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/929404
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/931836
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 09/941992
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: 09/946374
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 10/001054
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 10/081056
; PRIOR FILING DATE: 2002-02-20
; PRIOR APPLICATION NUMBER: 10/119480
; PRIOR FILING DATE: 2002-04-09

```

Query Match 1.3%; Score 7; DB 9; Length 489;

Best Local Similarity 100.0%; Pred. No. 4e+02; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 GVRVRL 133  
Db 353 GVRVRL 359

```

; RESULT 34
; US-10-238-283-42
; Sequence 42, Application US/10238283
; Publication No. US20030073190A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin

```

```

; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C15
; CURRENT APPLICATION NUMBER: US/10/238,283
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining prior application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 42
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-238-283-42

```

Query Match 1.3%; Score 7; DB 9; Length 489;

Best Local Similarity 100.0%; Pred. No. 4e+02; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 127 GVRVRL 133  
Db 353 GVRVRL 359

```

; RESULT 35
; US-10-238-370-42
; Sequence 42, Application US/10238370
; Publication No. US20030073191A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C10

```



CURRENT APPLICATION NUMBER: US/10/238,370  
CURRENT FILING DATE: 2002-09-09  
PRIOR APPLICATION NUMBER: 10/197942  
PRIOR FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/059114  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/065027  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/086478  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090689  
PRIOR FILING DATE: 1998-06-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 42  
LENGTH: 489  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-238-370-42

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 GVRVRL 133  
Db 353 GVRVRL 359

RESULT 36  
US-10-245-055-42  
Sequence 42, Application US/10245055  
Publication No. US20030073192A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Goddard, Audrey  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Phillippe  
APPLICANT: Watanabe, Colin  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
APPLICANT: Fong, Sherman  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3630R1C88  
CURRENT APPLICATION NUMBER: US/10/245,055  
CURRENT FILING DATE: 2002-09-16  
PRIOR APPLICATION NUMBER: 10/197942  
PRIOR FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/059114  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/065027  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/086478  
PRIOR FILING DATE: 1998-05-22

PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090689  
PRIOR FILING DATE: 1998-06-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 42  
LENGTH: 489  
TYPE: PRT  
ORGANISM: Homo Sapien  
US-10-245-055-42

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 127 GVRVRL 133  
Db 353 GVRVRL 359

RESULT 37  
US-10-245-147-42  
Sequence 42, Application US/10245147  
Publication No. US20030073193A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Goddard, Audrey  
APPLICANT: Grimaldi, J. Christopher  
APPLICANT: Gurney, Austin  
APPLICANT: Smith, Victoria  
APPLICANT: Stephan, Jean-Phillippe  
APPLICANT: Watanabe, Colin  
APPLICANT: Wood, William  
APPLICANT: Zhang, Zemin  
APPLICANT: Fong, Sherman  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
FILE REFERENCE: P3630R1C72  
CURRENT APPLICATION NUMBER: US/10/245,147  
CURRENT FILING DATE: 2002-09-16  
PRIOR APPLICATION NUMBER: 10/197942  
PRIOR FILING DATE: 2002-07-18  
PRIOR APPLICATION NUMBER: 60/059114  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/063046  
PRIOR FILING DATE: 1997-10-24  
PRIOR APPLICATION NUMBER: 60/065027  
PRIOR FILING DATE: 1997-11-10  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/086478  
PRIOR FILING DATE: 1998-05-22  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/089801  
PRIOR FILING DATE: 1998-06-18  
PRIOR APPLICATION NUMBER: 60/090557  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: 60/090689  
PRIOR FILING DATE: 1998-06-25  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 116  
SEQ ID NO 42  
LENGTH: 489  
TYPE: PRT  
ORGANISM: Homo Sapien

US-10-245-147-42

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GVRVRL 133  
|||||  
DB 353 GVRVRL 359

RESULT 38

US-10-245-730-42  
; Sequence 42, Application US/10245730  
; Publication No. US20030073194A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Baton, Dan  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Phillippe  
; APPLICANT: Matambe, Colin  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Fong, Sherman  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3630R1C85  
; CURRENT APPLICATION NUMBER: US/10/245,730  
; PRIOR FILING DATE: 2002-09-16  
; PRIOR APPLICATION NUMBER: 10/197942  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/059114  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/063046  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/065027  
; PRIOR FILING DATE: 1997-11-10  
; PRIOR APPLICATION NUMBER: 60/079689  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/086478  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090689  
; PRIOR FILING DATE: 1998-06-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 116  
; SEQ ID NO 42  
; LENGTH: 489  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-245-730-42

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GVRVRL 133  
|||||  
DB 353 GVRVRL 359

RESULT 39  
US-10-245-739-42  
; Sequence 42, Application US/10245739

Publication No. US20030073195A1

; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Baton, Dan  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Phillippe  
; APPLICANT: Matambe, Colin  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Fong, Sherman  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3630R1C86  
; CURRENT APPLICATION NUMBER: US/10/245,739  
; PRIOR FILING DATE: 2002-09-16  
; PRIOR APPLICATION NUMBER: 10/197942  
; PRIOR FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/059114  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/063046  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: 60/065027  
; PRIOR FILING DATE: 1997-11-10  
; PRIOR APPLICATION NUMBER: 60/079689  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: 60/086478  
; PRIOR FILING DATE: 1998-05-22  
; PRIOR APPLICATION NUMBER: 60/087607  
; PRIOR FILING DATE: 1998-06-02  
; PRIOR APPLICATION NUMBER: 60/089801  
; PRIOR FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: 60/090557  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 60/090689  
; PRIOR FILING DATE: 1998-06-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 116  
; SEQ ID NO 42  
; LENGTH: 489  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
US-10-245-739-42

Query Match 1.3%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GVRVRL 133  
|||||  
DB 353 GVRVRL 359

RESULT 40  
US-10-246-210-42  
; Sequence 42, Application US/10246210  
; Publication No. US20030073196A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin  
; APPLICANT: Baton, Dan  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Grimaldi, J. Christopher  
; APPLICANT: Gurney, Austin  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stephan, Jean-Phillippe  
; APPLICANT: Matambe, Colin  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; APPLICANT: Fong, Sherman

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
;; FILE REFERENCE: P3630R1C121  
;; CURRENT APPLICATION NUMBER: US/10/246,210  
;; CURRENT FILING DATE: 2002-09-18  
;; PRIOR APPLICATION NUMBER: 10/197942  
;; PRIOR FILING DATE: 2002-07-18  
;; PRIOR APPLICATION NUMBER: 60/059114  
;; PRIOR FILING DATE: 1997-09-17  
;; PRIOR APPLICATION NUMBER: 60/063046  
;; PRIOR FILING DATE: 1997-10-24  
;; PRIOR APPLICATION NUMBER: 60/065027  
;; PRIOR FILING DATE: 1997-11-10  
;; PRIOR APPLICATION NUMBER: 60/079689  
;; PRIOR FILING DATE: 1998-03-27  
;; PRIOR APPLICATION NUMBER: 60/086478  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087607  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/083801  
;; PRIOR FILING DATE: 1998-06-18  
;; PRIOR APPLICATION NUMBER: 60/090557  
;; PRIOR FILING DATE: 1998-06-24  
;; PRIOR APPLICATION NUMBER: 60/090689  
;; PRIOR FILING DATE: 1998-06-25  
;; Remaining Prior Application data removed - See File Wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 116  
;; SEQ ID NO: 42  
;; LENGTH: 489  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-246-210-42

Query Match 1.3% Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 127 GVRVRL 133  
DB 353 GVRVRL 359

Search completed: May 12, 2003, 10:16:07  
Job time : 52 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: May 12, 2003, 10:05:16 ; Search time 21 Seconds  
(without alignments)  
2403.361 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 525  
Sequence: 1 MRANPKTQAMPSEFTISLMKT.....KLMKRIAKILSLPIESIL 525

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

1: PIR\_73:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
4: PIR4:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525	100.0	525	2	B81859 phopholipase D-fam
2	237	45.1	508	2	B81083 cardiolipin synthase
3	15	2.9	495	2	A80636 conserved hypothet
4	14	2.7	493	2	H80806 probable synthase
5	14	2.7	493	2	D85666 probable synthase
6	14	2.7	493	2	C64847 ymcC protein - Esc
7	11	2.1	518	2	G97556 hypothetical 55.9K
8	11	2.1	518	2	A82777 phospholipase D fa
9	9	1.7	378	2	P35403 probable polypreny
10	8	1.5	161	2	AC1951 hypothetical prote
11	8	1.5	252	2	E86950 conserved hypothet
12	8	1.5	576	2	S49376 hypothetical prote
13	8	1.5	590	2	D83063 hypothetical prote
14	8	1.5	652	2	B82724 cardiolipin syntha
15	8	1.5	662	2	G95929 probable methylcro
16	8	1.5	1275	2	D38164 COBN protein - Pse
17	8	1.5	2599	2	A86616 unknown protein FI
18	8	1.5	8243	2	T31307 type I fatty acid
19	8	1.3	71	2	A70693 hypothetical prote
20	7	1.3	77	2	S29563 endothelin 2 precu
21	7	1.3	115	2	S09751 hypothetical prote
22	7	1.3	117	2	B83964 hypothetical prote
23	7	1.3	125	2	G72716 hypothetical prote
24	7	1.3	127	2	T31900 hypothetical prote
25	7	1.3	136	2	F64712 penicillin-binding
26	7	1.3	136	2	D71806 hypothetical prote
27	7	1.3	148	2	S68817 legumin beta chain
28	7	1.3	148	2	S68818 legumin beta chain
29	7	1.3	148	2	E97502 riboflavin synthas

30	7	1.3	148	2	A12720 riboflavin synthas
31	7	1.3	157	2	AE3400 riboflavin synthas
32	7	1.3	159	2	C84290 hypothetical prote
33	7	1.3	161	2	T48767 hypothetical prote
34	7	1.3	167	2	F69992 thioredoxin peroxi
35	7	1.3	172	2	AF0803 probable lipoprote
36	7	1.3	178	1	A39070 endothelin 2 precu
37	7	1.3	178	2	T30717 probable virion pr
38	7	1.3	178	2	T32796 hypothetical prote
39	7	1.3	181	1	B64167 cytochrome c bioge
40	7	1.3	188	2	AB3472 hypothetical prote
41	7	1.3	192	2	B71887 phosphohexose iso
42	7	1.3	199	2	JC5126 polyU-preferential
43	7	1.3	199	2	T26949 hypothetical prote
44	7	1.3	211	1	D69765 lactam utilization
45	7	1.3	218	2	S71260 embryonic protein
46	7	1.3	218	2	E82938 conserved hypothet
47	7	1.3	221	2	AH0286 phage shock protei
48	7	1.3	222	2	A99864 phage shock protei
49	7	1.3	222	2	H82169 phage shock protei
50	7	1.3	222	2	AF0658 phage shock protei
51	7	1.3	222	2	H85754 phage shock protei
52	7	1.3	222	2	C64879 phage shock protei
53	7	1.3	223	2	H87335 conserved hypothet
54	7	1.3	238	2	H90219 conserved hypothet
55	7	1.3	239	2	S07377 ribosomal prot
56	7	1.3	240	2	T02059 ribosomal protein
57	7	1.3	243	2	C96909 GTP-binding regula
58	7	1.3	244	2	C97532 conserved membrane
59	7	1.3	244	2	AC2751 probable transcrip
60	7	1.3	253	1	R52PDA ribosomal protein
61	7	1.3	254	1	HLHDA MHC class II histo
62	7	1.3	254	2	S50243 ribosomal protein
63	7	1.3	260	2	T18676 hypothetical prote
64	7	1.3	260	2	F69077 hypothetical prote
65	7	1.3	263	2	D84226 hypothetical prote
66	7	1.3	276	2	AD3486 hypothetical prote
67	7	1.3	286	2	T16681 hypothetical prote
68	7	1.3	291	1	B53419 biphenyl-2,3-diol
69	7	1.3	293	2	C71283 hypothetical prote
70	7	1.3	297	2	H72530 hypothetical prote
71	7	1.3	311	2	T40921 hypothetical prote
72	7	1.3	311	2	AF2353 hypothetical prote
73	7	1.3	317	2	D75296 signal recognition
74	7	1.3	318	2	C70636 probable echA3 pr
75	7	1.3	319	2	E82178 homoserine O-succi
76	7	1.3	319	2	T43040 hypothetical prote
77	7	1.3	323	1	F64360 conserved hypothet
78	7	1.3	323	2	T44256 thiamin biosynthes
79	7	1.3	327	2	A89828 mevalonate diphosp
80	7	1.3	329	2	A69791 hypothetical prote
81	7	1.3	337	2	F75599 conserved hypothet
82	7	1.3	354	2	I46863 macropage scaveng
83	7	1.3	354	2	A82544 translaton initia
84	7	1.3	357	2	E85092 hypothetical prote
85	7	1.3	367	2	F69816 reticuline oxidase
86	7	1.3	370	2	A96547 hypothetical prote
87	7	1.3	371	2	T21707 hypothetical prote
88	7	1.3	374	2	T22149 hypothetical prote
89	7	1.3	375	2	T03256 GTP-binding protei
90	7	1.3	375	2	T52300 H+-exporting ATPas
91	7	1.3	377	2	T16985 GTP-binding protei
92	7	1.3	377	2	T07376 GTP-binding protei
93	7	1.3	377	2	T04086 GTP-binding protei
94	7	1.3	377	2	T04089 GTP-binding protei
95	7	1.3	377	2	T02084 G-box binding fact
96	7	1.3	380	2	T02085 GTP-binding protei
97	7	1.3	380	2	T03765 GTP-binding regula
98	7	1.3	380	2	T08036 GTP-binding regula
99	7	1.3	380	2	JN0823 schA protein - Str
100	7	1.3	389	2	S74554 hypothetical prote
101	7	1.3	396	2	A71296 D-alanine-D-alanin
102	7	1.3	397	1	XURT acetyl-CoA C-acyl

103	7	1.3	403	2	T20727	hypothetical prote
104	7	1.3	412	2	T23385	hypothetical prote
105	7	1.3	414	2	AF0389	exonuclease sbcd. l
106	7	1.3	418	2	G84546	probable tetracycl
107	7	1.3	419	2	H96601	hypothetical prote
108	7	1.3	422	1	A60503	sperm-binding glyc
109	7	1.3	432	2	AT1932	two-component resp
110	7	1.3	436	2	AA1229	transcription init
111	7	1.3	443	2	T35776	hypothetical prote
112	7	1.3	447	2	T35519	hypothetical prote
113	7	1.3	454	2	AA4407	macrophage scaveng
114	7	1.3	454	2	I46862	macrophage scaveng
115	7	1.3	457	2	C65203	argininosuccinate
116	7	1.3	457	2	A88240	argininosuccinate
117	7	1.3	457	2	R60807	argininosuccinate
118	7	1.3	458	2	C69421	methylylologen-red
119	7	1.3	465	2	AH0275	funarate hydratase
120	7	1.3	467	2	S16915	gamma-aminobutyric
121	7	1.3	467	2	S19317	gamma-aminobutyric
122	7	1.3	474	1	G2MS11	Ig gamma-2b chain
123	7	1.3	477	2	T05202	pectinesterase hom
124	7	1.3	491	2	C75078	hypothetical prote
125	7	1.3	492	2	T14676	hypothetical prote
126	7	1.3	494	2	S62902	legumin 2 precursor
127	7	1.3	500	2	I56502	4-aminobutyrate tr
128	7	1.3	502	2	F64543	conserved hypothet
129	7	1.3	502	2	E71963	probable cardiolip
130	7	1.3	504	2	AT1075	cardiolipin syntha
131	7	1.3	504	2	AT1433	cardiolipin syntha
132	7	1.3	510	2	S62901	legumin 3i precurs
133	7	1.3	510	2	T26535	hypothetical prote
134	7	1.3	512	2	G91200	type III secretion
135	7	1.3	512	2	I80311	sepc protein - fsc
136	7	1.3	512	2	C66047	escsc (imported) -
137	7	1.3	513	2	B96524	hypothetical prote
138	7	1.3	516	2	G70556	probable tryp prote
139	7	1.3	519	2	T14938	hypothetical prote
140	7	1.3	528	1	WMBV8B	58k protein - bart
141	7	1.3	528	2	G70854	probable sera prot
142	7	1.3	528	2	T45418	phosphoglycerate d
143	7	1.3	529	2	T45254	probable anthranil
144	7	1.3	529	2	F82983	conserved hypothet
145	7	1.3	533	2	G75348	conserved hypothet
146	7	1.3	542	2	E86628	prophage psi prote
147	7	1.3	556	2	B64482	adenine deaminase
148	7	1.3	565	2	T47423	hypothetical prote
149	7	1.3	566	2	AT0164	phenylalanine-trna
150	7	1.3	574	2	G96572	protein f12M16.12
151	7	1.3	587	2	A90394	bps2 protein homol
152	7	1.3	602	2	G81654	conserved hypothet
153	7	1.3	621	2	H87662	feruloyl-coa synth
154	7	1.3	643	2	E70682	probable gamma-glu
155	7	1.3	657	2	A60431	probable exported
156	7	1.3	682	2	JC7385	multispecific orga
157	7	1.3	686	2	H90798	Iha adhesin (import
158	7	1.3	696	2	G85607	probable receptor
159	7	1.3	701	2	AE2795	ATP-dependent DNA
160	7	1.3	704	2	E97574	ATP-dependent DNA
161	7	1.3	773	2	T46010	anthranilate phosph
162	7	1.3	776	2	C96554	unknown protein (i
163	7	1.3	805	2	I40029	nitrite reductase
164	7	1.3	811	2	AA1054	fasciclin II, tran
165	7	1.3	830	2	S36018	gene transtrack-p88
166	7	1.3	830	2	A84951	DNA topoisomerase
167	7	1.3	873	2	B41054	fasciclin II PI-1i
168	7	1.3	876	2	T01117	hypothetical prote
169	7	1.3	953	2	T40643	probable serine th
170	7	1.3	974	2	T04910	hypothetical prote
171	7	1.3	990	2	S67499	glutamate synthase
172	7	1.3	1021	2	I39207	leukocyte surface
173	7	1.3	1133	2	T01920	probable RNA-direc
174	7	1.3	1173	2	I50620	prockr2 - chicken
175	7	1.3	1189	2	AH2154	WD-repeat protein
176	7	1.3	1270	2	A96669	protein FIN19.11 l
177	7	1.3	1271	2	T08607	hypothetical prote
178	7	1.3	1317	2	F83310	conserved hypothet
179	7	1.3	1414	2	B70674	probable mbt prote
180	7	1.3	1463	2	T30290	AAS surface protei
181	7	1.3	1495	2	S22610	DNA (cytosine-5'-)
182	7	1.3	1498	2	AF1082	B. subtilis yuka p
183	7	1.3	1498	2	AG1439	B. subtilis yuka p
184	7	1.3	1537	2	JC4172	DNA (cytosine-5'-)
185	7	1.3	1999	2	AB2018	hypothetical prote
186	7	1.3	2708	2	T09079	probable chloroqui
187	7	1.3	2819	2	T09080	probable chloroqui
188	7	1.3	3069	2	H70656	lactyl-acid synthas
189	7	1.3	3076	2	AB7058	probable autotrans
190	7	1.3	3705	2	AD0123	mycosubtilin synth
191	7	1.3	5369	2	T44807	probable trp opero
192	6	1.1	20	2	AC0269	hypothetical prote
193	6	1.1	33	2	C95200	hypothetical prote
194	6	1.1	39	2	B70207	hypothetical prote
195	6	1.1	42	1	Q06C45	hypothetical prote
196	6	1.1	42	1	T04383	zein protein - mai
197	6	1.1	48	1	Q1BP87	gene 1.8 protein -
198	6	1.1	48	2	AB3386	hypothetical prote
199	6	1.1	51	2	S51909	hypothetical prote
200	6	1.1	52	2	DA3259	cryptogene protein
201	6	1.1	53	2	F70956	H+-transporting tw
202	6	1.1	54	2	B90538	hypothetical prote
203	6	1.1	55	2	H81732	hypothetical prote
204	6	1.1	60	2	B28918	hypothetical prote
205	6	1.1	61	2	S78741	protein YPR170w-a
206	6	1.1	63	2	C84377	hypothetical prote
207	6	1.1	65	2	A75543	transcription regu
208	6	1.1	65	2	S08014	calcium-binding pr
209	6	1.1	65	2	T12900	hypothetical prote
210	6	1.1	68	1	JOBP1	head-to-tail joini
211	6	1.1	68	2	T13089	probable head-to-t
212	6	1.1	68	2	G90832	head-to-tail joini
213	6	1.1	68	2	B90901	ribosomal protein
214	6	1.1	70	2	S73177	hypothetical prote
215	6	1.1	72	2	C83219	hypothetical prote
216	6	1.1	73	2	H98129	hypothetical prote
217	6	1.1	73	2	AG3157	hypothetical prote
218	6	1.1	73	2	C65069	hypothetical prote
219	6	1.1	74	2	E71341	hypothetical prote
220	6	1.1	75	2	A60698	probable pathogeni
221	6	1.1	75	2	AG2371	hypothetical prote
222	6	1.1	76	2	AI2300	integrin alpha 5 s
223	6	1.1	77	2	E87077	hypothetical prote
224	6	1.1	78	1	A05008	conserved hypothet
225	6	1.1	83	2	UC7607	translacion initia
226	6	1.1	83	2	T42102	prolactin-releasin
227	6	1.1	83	2	B75418	hypothetical prote
228	6	1.1	85	2	T40431	hypothetical prote
229	6	1.1	85	2	G87357	conserved hypothet
230	6	1.1	88	2	F97400	hypothetical prote
231	6	1.1	88	2	AH0760	probable propanedi
232	6	1.1	91	2	AE0368	hypothetical prote
233	6	1.1	91	2	B75602	ribosomal protein
234	6	1.1	92	2	S78386	hypothetical prote
235	6	1.1	92	2	DB1731	conserved hypothet
236	6	1.1	92	2	H82867	transcription regu
237	6	1.1	94	2	S22343	chaperonin groes -
238	6	1.1	94	2	S59080	hypothetical prote
239	6	1.1	95	2	B64133	monocytic cytokine
240	6	1.1	95	2	AB48093	prophage psi prote
241	6	1.1	97	2	E86866	hypothetical prote
242	6	1.1	98	2	S77121	hypothetical prote
243	6	1.1	98	2	A59286	monocyte chemotact
244	6	1.1	99	1	JC5295	monocyte chemotact
245	6	1.1	99	2	JC2336	monocyte chemotact
246	6	1.1	99	2	JC2136	monocyte chemotact
247	6	1.1	99	2	A60299	monocyte chemotact
248	6	1.1	99	2		

249	6	1.1	99	2	JC2417	monocyte chemotatr	322	1.1	141	2	A27249	vitelline membrane
250	6	1.1	99	2	F82709	hypothetical prote	323	1.1	142	2	H71061	hypothetical prote
251	6	1.1	100	2	T07466	ribosomal protein	324	1.1	142	2	B83381	hypothetical prote
252	6	1.1	100	2	T49857	hypothetical prote	325	1.1	143	2	S20937	photosystem I chal
253	6	1.1	100	2	AE1654	hypothetical prote	326	1.1	143	2	S05717	hypothetical prote
254	6	1.1	100	2	AI1282	hypothetical prote	327	1.1	144	2	AB0325	urease (EC 3.5.1.5
255	6	1.1	101	1	R3N718	ribosomal protein	328	1.1	144	2	T35720	hypothetical, prote
256	6	1.1	102	2	T49651	hypothetical prote	329	1.1	145	2	AI1071	DNA polymerase III
257	6	1.1	102	2	B95336	hypothetical prote	330	1.1	145	2	A95015	hypothetical prote
258	6	1.1	103	2	S20763	nonstructural prot	331	1.1	145	2	A86854	conserved hypotnet
259	6	1.1	103	2	H70956	hypothetical prote	332	1.1	145	2	D72726	probable ribosomal
260	6	1.1	103	2	AF0031	probable membrane	333	1.1	146	2	T14667	chemotaxis respons
261	6	1.1	104	1	CCMKP	cytochrome c - spi	334	1.1	146	2	T18096	hypothetical prote
262	6	1.1	104	2	AF1878	hypothetical prote	335	1.1	146	2	T15742	hypothetical prote
263	6	1.1	104	2	H87548	hypothetical prote	336	1.1	147	2	T20563	hemoglobin epsilon
264	6	1.1	106	2	S26464	hypothetical prote	337	1.1	147	2	F64715	hypothetical prote
265	6	1.1	107	2	B83899	Ig heavy chain V r	338	1.1	147	2	T40625	hypothetical prote
266	6	1.1	108	2	A49831	mutants block spor	339	1.1	147	2	S68815	legumin beta chain
267	6	1.1	108	2	S34940	nitrogenase (Ec 1,	340	1.1	148	2	D86365	hypothetical prote
268	6	1.1	108	2	S34942	fimbrial protein S	341	1.1	148	2	F95170	hypothetical prote
269	6	1.1	108	2	S69294	hypothetical prote	342	1.1	148	2	C87300	conserved hypotnet
270	6	1.1	108	2	AC1857	hypothetical prote	343	1.1	148	2	R5EC9	ribosomal protein
271	6	1.1	109	2	AS4678	monocyte chemotact	344	1.1	149	1	C91276	50S ribosomal subu
272	6	1.1	110	2	S61540	ribosomal protein	345	1.1	149	2	C86117	50S ribosomal subu
273	6	1.1	110	2	T26817	hypothetical prote	346	1.1	149	2	AB1053	50S ribosomal chal
274	6	1.1	112	2	S73078	hypothetical prote	347	1.1	149	2	S49425	sulfide dehydrogen
275	6	1.1	113	2	D89762	hypothetical prote	348	1.1	149	2	T05965	probable pyrroline
276	6	1.1	113	2	F72692	hypothetical prote	349	1.1	150	2	AI0429	50S ribosomal) prot
277	6	1.1	114	1	W7W41	E7 protein - human	350	1.1	150	2	S37014	transposase (clone
278	6	1.1	114	2	S73778	MG267 Homolog F11-	351	1.1	150	2	E96521	protein F21D18.14
279	6	1.1	115	2	E64229	hypothetical prote	352	1.1	150	2	F98036	hypothetical prote
280	6	1.1	115	2	D95310	hypothetical prote	353	1.1	150	2	T50648	phloem protein (lm
281	6	1.1	115	2	E75451	conserved hypotnet	354	1.1	150	2	C72697	hypothetical prote
282	6	1.1	116	2	F64510	hypothetical prote	355	1.1	150	2	T14948	hypothetical prote
283	6	1.1	116	2	H75601	hypothetical prote	356	1.1	151	2	JC7325	allatostatin prepr
284	6	1.1	117	2	A87029	hypothetical prote	357	1.1	151	2	S29752	ubiquitin-activati
285	6	1.1	120	2	C86681	prophage pil prote	358	1.1	152	2	D95961	hypothetical prote
286	6	1.1	120	2	AI2380	hypothetical prote	359	1.1	152	2	C64373	hypothetical prote
287	6	1.1	122	2	AD2072	hypothetical prote	360	1.1	153	2	D86738	transcription regu
288	6	1.1	123	2	A87216	hypothetical prote	361	1.1	154	2	S53572	probable membrane
289	6	1.1	123	2	B87355	hypothetical prote	362	1.1	158	2	T19248	hypothetical prote
290	6	1.1	124	2	C69335	succinate dehydrog	363	1.1	158	2	F90958	conserved hypotnet
291	6	1.1	124	2	AG2441	hypothetical prote	364	1.1	158	2	AG0223	hypothetical prote
292	6	1.1	125	2	I46857	monocyte chemotatr	365	1.1	158	2	S76604	hypothetical prote
293	6	1.1	125	2	A70020	hypothetical prote	366	1.1	161	1	AE0981	probable acetyltra
294	6	1.1	125	2	D69798	hypothetical prote	367	1.1	161	2	C97451	hypothetical prote
295	6	1.1	127	2	F72561	hypothetical prote	368	1.1	161	2	A99935	spheroplast protel
296	6	1.1	128	2	C83085	conserved hypotnet	369	1.1	161	2	E85783	hypothetical prote
297	6	1.1	129	2	T44125	hypothetical prote	370	1.1	161	2	G64933	beta-lactoglobulin
298	6	1.1	130	2	T29155	hypothetical prote	371	1.1	161	2	LGHOD	hypothetical prote
299	6	1.1	131	2	T22997	hypothetical prote	372	1.1	162	1	T33127	homoproteococcat
300	6	1.1	131	2	E70920	probable moab prot	373	1.1	162	2	G70440	transcription elon
301	6	1.1	131	2	AG1421	hypothetical prote	374	1.1	163	2	AC3525	probable cellsurfa
302	6	1.1	133	2	JC2007	hypothetical prote	375	1.1	163	2	B64628	hypothetical prote
303	6	1.1	133	2	AB3397	differentiation in	376	1.1	164	2	AE1644	hypothetical prote
304	6	1.1	133	2	T44627	hypothetical prote	377	1.1	164	2	AF1281	probable cellsurfa
305	6	1.1	134	2	A45893	T-cell receptor al	378	1.1	164	2	B71138	hypothetical prote
306	6	1.1	134	2	S52117	hyperglycemic horm	379	1.1	165	2	D72085	hypothetical prote
307	6	1.1	134	2	A40227	transcription repr	380	1.1	165	2	E86537	hypothetical prote
308	6	1.1	134	2	JC2112	helix-loop-helix p	381	1.1	165	2	S66109	2-amino-4-hydroxy-
309	6	1.1	135	2	H82021	truncated pilln NM	382	1.1	167	1	E87085	arginine repressor
310	6	1.1	135	2	S48142	hypoglycemic hormo	383	1.1	167	2	T01155	hypothetical prote
311	6	1.1	135	2	S48141	hypoglycemic hormo	384	1.1	167	2	S59712	hypothetical prote
312	6	1.1	136	2	C72575	hypothetical prote	385	1.1	169	2	T35937	hypothetical prote
313	6	1.1	137	2	S52370	superoxide dismuta	386	1.1	169	2	T46706	Ig gamma-2a chain
314	6	1.1	137	2	C45893	T-cell receptor al	387	1.1	169	2	A35944	hypothetical prote
315	6	1.1	137	2	H82537	Tonb protein XR260	388	1.1	170	2	G70638	hypothetical prote
316	6	1.1	139	2	T33713	conserved hypotnet	389	1.1	170	2	AE1817	hypothetical prote
317	6	1.1	139	2	AB3604	sugar-binding prot	390	1.1	170	2	S15911	hypothetical prote
318	6	1.1	140	2	S26384	hypothetical prote	391	1.1	171	2	F70789	hypothetical prote
319	6	1.1	141	1	HAIG1	hemoglobin alpha-1	392	1.1	172	2	H69936	hypothetical prote
320	6	1.1	141	1	E32998	chorion protein SI	393	1.1	172	2		
321	6	1.1	141	2	D69160	hypothetical prote	394	1.1	172	2		





541	6	1.1	218	2	S07359	614	1.1	242	2	I64025	hypothetical prote
542	6	1.1	218	2	S0287	615	1.1	242	2	AD1524	B subtilis vYpB p
543	6	1.1	218	2	G82189	616	1.1	242	2	G97573	fine protein VCA01
544	6	1.1	218	2	AE1147	617	1.1	242	2	AH2794	polyketide biosynt
545	6	1.1	219	2	T42605	618	1.1	243	2	F82497	uridine phosphoryl
546	6	1.1	219	2	AE1506	619	1.1	243	2	A95064	conserved hypochet
547	6	1.1	220	2	F97413	620	1.1	243	2	T18851	hypothetical prote
548	6	1.1	220	2	AD2631	621	1.1	243	2	AG3465	oxidoreductase ucp
549	6	1.1	221	2	S76798	622	1.1	244	2	S31864	chlorophyll a/b-bi
550	6	1.1	221	2	T52138	623	1.1	244	2	T26704	hypothetical prote
551	6	1.1	222	2	E71507	624	1.1	245	2	C70090	hypothetical prote
552	6	1.1	222	2	C81661	625	1.1	245	2	E86407	hypothetical prote
553	6	1.1	222	2	H69118	626	1.1	245	2	S38950	Ig gamma chain - m
554	6	1.1	223	2	AC2871	627	1.1	246	2	E90216	ABC transporter, A
555	6	1.1	223	2	E97647	628	1.1	246	2	AG2988	competence protein
556	6	1.1	223	2	A86057	629	1.1	246	2	E81305	probable membrane
557	6	1.1	223	2	F91210	630	1.1	246	2	D70969	hypothetical prote
558	6	1.1	224	2	T51742	631	1.1	246	2	B82764	hypothetical prote
559	6	1.1	224	2	D71511	632	1.1	246	2	E64393	hypothetical prote
560	6	1.1	224	2	C97750	633	1.1	247	2	T45502	hypothetical prote
561	6	1.1	225	1	MMN29	634	1.1	247	2	S77511	hypothetical prote
562	6	1.1	225	1	T21655	635	1.1	247	2	T21406	hypothetical prote
563	6	1.1	225	2	E69896	636	1.1	247	2	A86353	protein F2E2.4 (im
564	6	1.1	226	2	B69771	637	1.1	247	2	H83176	probable glycosyl
565	6	1.1	226	2	T36435	638	1.1	248	1	S22853	probable response
566	6	1.1	227	2	I61849	639	1.1	248	2	G83253	glycosyl transfera
567	6	1.1	227	2	I36932	640	1.1	248	2	S30194	ribosomal protein
568	6	1.1	227	2	E87118	641	1.1	248	2	B83030	probable rRNA meth
569	6	1.1	227	2	G97112	642	1.1	248	2	C90984	probable glycosyl
570	6	1.1	227	2	T21662	643	1.1	248	2	S57910	probable response
571	6	1.1	227	2	T35395	644	1.1	249	1	R3H06	ribosomal protein
572	6	1.1	228	2	G64214	645	1.1	249	1	R3MS6	ribosomal protein
573	6	1.1	228	2	AD3458	646	1.1	249	1	R3MS6	probable dehydrat
574	6	1.1	229	2	AB1106	647	1.1	249	2	B96022	phosphoglyceromuta
575	6	1.1	229	2	AB1468	648	1.1	249	2	B82625	ribosomal protein
576	6	1.1	229	2	A64137	649	1.1	249	2	UC4145	ribosomal protein
577	6	1.1	229	2	D81867	650	1.1	249	2	T50003	40S ribosomal prote
578	6	1.1	229	2	H81073	651	1.1	249	2	S18677	ATP-dependent RNA
579	6	1.1	229	2	T15359	652	1.1	249	2	S35491	rnp30 protein - ra
580	6	1.1	229	2	AD1193	653	1.1	249	2	E82638	hypothetical prote
581	6	1.1	230	2	T37016	654	1.1	249	2	T05424	hypothetical prote
582	6	1.1	231	2	PC4155	655	1.1	249	2	D95327	probable response
583	6	1.1	231	2	H70407	656	1.1	249	2	T05100	ribosomal protein
584	6	1.1	231	2	AI3198	657	1.1	250	2	E97385	hypothetical prote
585	6	1.1	231	2	AH3406	658	1.1	250	2	A62603	components of type
586	6	1.1	232	2	G96702	659	1.1	250	2	T04334	ribosomal protein
587	6	1.1	233	2	C90160	660	1.1	250	2	S31863	transcription regu
588	6	1.1	233	2	T03802	661	1.1	250	2	A83844	4-amin-4-deoxycho
589	6	1.1	233	2	B95278	662	1.1	251	2	F87614	Al3 antigen - Tryp
590	6	1.1	234	2	S76694	663	1.1	251	2	E73491	fine protein - Del
591	6	1.1	234	2	S06141	664	1.1	251	2	T50432	conserved hypochet
592	6	1.1	234	2	S77098	665	1.1	251	2	T32086	hypothetical prote
593	6	1.1	235	2	T14408	666	1.1	251	2	A44844	phosphoprotein pho
594	6	1.1	236	1	E65073	667	1.1	252	2	A65174	hypothetical 28.0
595	6	1.1	236	1	D86778	668	1.1	252	2	AG1210	hypothetical prote
596	6	1.1	236	2	E91099	669	1.1	253	2	D86056	hypothetical prote
597	6	1.1	236	2	A85945	670	1.1	253	2	T30381	hypothetical prote
598	6	1.1	236	2	B98315	671	1.1	253	2	JC5347	cdd3 protein - Clo
599	6	1.1	236	2	AI2967	672	1.1	253	2	AF1159	flagellar biosynth
600	6	1.1	236	2	B86881	673	1.1	253	2	AF1518	flagellar biosynth
601	6	1.1	236	2	H85435	674	1.1	253	2	B97465	hypothetical prote
602	6	1.1	237	2	H64461	675	1.1	253	2	S35585	chymotrypsin-like
603	6	1.1	237	2	T14409	676	1.1	254	2	B31790	Ig heavy chain V r
604	6	1.1	237	2	F97109	677	1.1	254	2	S73385	hypothetical prote
605	6	1.1	237	2	G70153	678	1.1	255	1	H64746	yaf1 protein - Esc
606	6	1.1	237	2	AH2570	679	1.1	255	2	B90660	probable antioctran
607	6	1.1	238	2	I40703	680	1.1	255	2	H85510	hypothetical prote
608	6	1.1	238	2	I62385	681	1.1	255	2	T45772	cytochrome P450 ho
609	6	1.1	238	2	T02373	682	1.1	255	2		
610	6	1.1	238	2	S49911	683	1.1	255	2		
611	6	1.1	239	1	R3ZP6E	684	1.1	255	2		
612	6	1.1	240	2	B87325	685	1.1	255	2		
613	6	1.1	240	2	T45727	686	1.1	255	2		



833	6	1.1	293	2	D81204	906	6	1.1	306	2	T25954	hypothetical prote
834	6	1.1	293	2	A81780	907	6	1.1	306	2	C97573	hypothetical prote
835	6	1.1	293	2	E81661	908	6	1.1	306	2	AD2794	conserved hypother
836	6	1.1	293	2	I39620	909	6	1.1	306	2	S50689	hypothetical prote
837	6	1.1	293	2	B97194	910	6	1.1	307	1	B96722	hypothetical prote
838	6	1.1	294	2	I39610	911	6	1.1	307	2	T24832	phosphoprotein pho
839	6	1.1	294	2	H64632	912	6	1.1	307	2	T34973	hypothetical prote
840	6	1.1	294	2	AE1271	913	6	1.1	307	2	H90237	5,10-methylenetet
841	6	1.1	294	2	AG1633	914	6	1.1	307	2	AD2575	ornithine carbanoy
842	6	1.1	295	2	S50135	915	6	1.1	308	2	D83452	chromosome partiti
843	6	1.1	295	2	AC1988	916	6	1.1	308	2	AB2833	probable cytochrom
844	6	1.1	295	2	S33663	917	6	1.1	308	2	AB3852	D-alanine-D-alanin
845	6	1.1	295	2	A97494	918	6	1.1	308	2	AE1610	pollen allergen pi
846	6	1.1	295	2	AC2712	919	6	1.1	308	2	AI1247	hypothetical prote
847	6	1.1	295	2	AC2858	920	6	1.1	308	2	E87380	hypothetical prote
848	6	1.1	296	1	J02155	921	6	1.1	308	2	G75264	2-dehydro-3-deoxy
849	6	1.1	296	1	T45203	922	6	1.1	308	2	G69135	probable lipase -
850	6	1.1	296	2	S44972	923	6	1.1	308	2	AE3182	hypothetical prote
851	6	1.1	296	2	B39581	924	6	1.1	309	2	T00503	transcription regu
852	6	1.1	296	2	JT0548	925	6	1.1	309	2	B69473	probable MYB faml
853	6	1.1	296	2	H98338	926	6	1.1	309	2	E86937	conserved hypother
854	6	1.1	296	2	AB2944	927	6	1.1	309	2	AB3809	conserved hypother
855	6	1.1	297	2	AC1915	928	6	1.1	310	2	G96761	protease homolog
856	6	1.1	297	2	H75449	929	6	1.1	310	2	A95287	probable MAP Kinas
857	6	1.1	297	2	A81002	930	6	1.1	310	2	D75202	probable dioxygena
858	6	1.1	297	2	H83058	931	6	1.1	310	2	A55768	dipeptide abc tran
859	6	1.1	297	2	F97999	932	6	1.1	310	2	T21375	asparaginyl-peptid
860	6	1.1	298	2	F81325	933	6	1.1	311	2	T43947	hypothetical prote
861	6	1.1	298	2	H82678	934	6	1.1	311	2	AE3527	N-acetyl-gamma-glu
862	6	1.1	298	2	T33317	935	6	1.1	311	2	AB1703	ornithine carbanoy
863	6	1.1	299	1	F88922	936	6	1.1	312	2	C72547	phospholipase D fa
864	6	1.1	299	1	NIA1F	937	6	1.1	312	2	B37245	probable aspartate
865	6	1.1	299	2	F84785	938	6	1.1	312	2	D91011	1-phosphofructokin
866	6	1.1	299	2	E87262	939	6	1.1	312	2	F85855	fructose-1-phospha
867	6	1.1	299	2	D95382	940	6	1.1	312	2	AE0158	1-phosphofructokin
868	6	1.1	300	2	S51044	941	6	1.1	312	2	AE0783	1-phosphofructokin
869	6	1.1	300	2	AS5792	942	6	1.1	312	2	H70433	1-phospholysaccharide
870	6	1.1	300	2	D87497	943	6	1.1	312	2	B81780	probable cytoplasm
871	6	1.1	300	2	A82017	944	6	1.1	312	2	C81204	conserved hypother
872	6	1.1	300	2	C71569	945	6	1.1	312	2	F86876	hypothetical prote
873	6	1.1	300	2	H89766	946	6	1.1	313	2	C71545	hypothetical prote
874	6	1.1	301	2	T52549	947	6	1.1	313	2	AE2812	probable oligopept
875	6	1.1	301	2	AE3465	948	6	1.1	313	2	T27221	conserved hypother
876	6	1.1	301	2	CE4011	949	6	1.1	313	2	A95268	hypothetical prote
877	6	1.1	302	1	H64127	950	6	1.1	313	2	AC2595	probable transcript
878	6	1.1	302	2	T36030	951	6	1.1	314	1	BMEN27	ABC transporter, s
879	6	1.1	302	2	G70614	952	6	1.1	314	2	F81699	peptide ABC transp
880	6	1.1	302	2	D65056	953	6	1.1	314	2	D72767	hypothetical prote
881	6	1.1	302	2	F91079	954	6	1.1	314	2	G68348	hypothetical prote
882	6	1.1	302	2	G85924	955	6	1.1	314	2	D90349	phosphate permease
883	6	1.1	302	2	AB0857	956	6	1.1	314	2	D70927	microsomal dipepti
884	6	1.1	302	2	D85253	957	6	1.1	314	2	S70101	probable lpw prot
885	6	1.1	302	2	F84806	958	6	1.1	315	2	AE3324	hypothetical prote
886	6	1.1	303	2	H71027	959	6	1.1	315	2	S22316	D-alanine-D-alanin
887	6	1.1	304	2	H88986	960	6	1.1	315	2	C87293	finger protein zip
888	6	1.1	304	2	H82984	961	6	1.1	316	2	A05043	hypothetical prote
889	6	1.1	304	2	H95346	962	6	1.1	316	2	AE2101	finger protein zip
890	6	1.1	305	2	S37086	963	6	1.1	316	2	E95398	acetyl-CoA carboxy
891	6	1.1	305	2	S55666	964	6	1.1	316	2	B69315	probable ABC trans
892	6	1.1	305	2	D95128	965	6	1.1	316	2	E88973	acetylornithine de
893	6	1.1	305	2	A97999	966	6	1.1	317	1	A42431	hypothetical prote
894	6	1.1	305	2	D69199	967	6	1.1	317	2	AB8672	3-oxoacyl-[acyl]-ca
895	6	1.1	305	2	F70533	968	6	1.1	317	2	E90812	3-oxoacyl-[acyl]-ca
896	6	1.1	305	2	C64864	969	6	1.1	317	2	AB0642	3-oxoacyl-[acyl]-ca
897	6	1.1	305	2	E90838	970	6	1.1	317	2	AC2209	4-diphosphocytidyl
898	6	1.1	305	2	E85696	971	6	1.1	317	2	T08962	hypothetical prote
899	6	1.1	305	2	AE0673	972	6	1.1	317	2	B70758	hypothetical prote
900	6	1.1	305	2	T06020	973	6	1.1	318	2	H96535	hypothetical prote
901	6	1.1	305	2	H95986	974	6	1.1	318	2	C83555	hypothetical prote
902	6	1.1	306	1	H65033	975	6	1.1	319	2	A70959	probable aspartate
903	6	1.1	306	2	F72606	976	6	1.1	319	2	T14831	acetyl-CoA carboxy
904	6	1.1	306	2	C91057	977	6	1.1	319	2	T15137	hypothetical prote
905	6	1.1	306	2	G85901	978	6	1.1	320	2	F84965	6-phosphofructokin

979	6	1.1	320	2	AF3237	transcriptional re
980	6	1.1	321	2	T07489	acetyl-CoA carboxy
981	6	1.1	321	2	B82259	lipidic acid synthase
982	6	1.1	322	2	S38194	ribosomal protein
983	6	1.1	322	2	G75291	conserved hypotet
984	6	1.1	322	2	D84567	hypothetical prote
985	6	1.1	322	2	A45036	single-stranded-DN
986	6	1.1	322	2	H83851	hypothetical prote
987	6	1.1	323	2	F97503	3-oxoacyl-lacyl-ca
988	6	1.1	323	2	A12721	3-oxoacyl-lacyl-ca
989	6	1.1	324	2	F75029	o-sialoglycoprotei
990	6	1.1	325	2	H72329	phenylalanine-tRNA
991	6	1.1	325	2	T02117	hypothetical prote
992	6	1.1	325	2	T15047	hypothetical prote
993	6	1.1	325	2	C97603	hypothetical prote
994	6	1.1	325	2	T06214	probable b-keto ac
995	6	1.1	326	1	J01238	zinc finger protei
996	6	1.1	326	2	T00640	peroxidase homolo
997	6	1.1	327	2	C97677	probable glycerol-
998	6	1.1	327	2	A12901	glycerol-3-phospha
999	6	1.1	327	2	H83145	lipate synthase P
1000	6	1.1	327	2	G97859	hypothetical prote

## ALIGNMENTS

**RESULT 1**

B81859  
phospholipase D-family protein NMA1646 [Imported] - Neisseria meningitidis (strain Z2491)

C.Species: Neisseria meningitidis

C.Date: 05-May-2000 #sequence-revision 05-May-2000 #text-change 02-Feb-2001

C.Accession: B81859

R.Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morello, R.; Holtroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A.Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A.Reference number: AB1775; MUID:2022556; PMID:10761919

A.Accession: B81859

A.Status: Preliminary

A.Molecule type: DNA

A.Residues: 1-525 <PAR>

A.Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84874.1; PTD:g7380202

A.Experimental source: serogroup A, strain Z2491

C.Genetics:

C.Gene: NMA1646

	Query Match	100.0%	Score 525:	DB 2:	Length 525:	
	Best Local Similarity	100.0%	Pred. No. 0:			
	Matches 525:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0:	
QY	1	MRANKTQAMSEETSLMKTRSLISLCLLCSCSGSWLPLEERTESHFNPTSKRVRLDN	60			
Db	1	MRANKTQAMSEETSLMKTRSLISLCLLCSCSGSWLPLEERTESHFNPTSKRVRLDN	60			
QY	61	ILQIHTHTHTNGSLDIYLLNDPHEAFARAAALIESAEHSLDQYITWNDISGRLLFMIV	120			
Db	61	ILQIHTHTHTNGSLDIYLLNDPHEAFARAAALIESAEHSLDQYITWNDISGRLLFMIV	120			
QY	121	YLAERGVRRVLLDDNNTTSGLDLLALLDSHPNIEVRLFNPFLVKRRALGYLLDFPRL	180			
Db	121	YLAERGVRRVLLDDNNTTSGLDLLALLDSHPNIEVRLFNPFLVKRRALGYLLDFPRL	180			
QY	181	NRRMHNSFTADNRRATILIGRNIGDEYFKVEDTYFADLLILATGSGVYGEVSHDFDRYWA	240			
Db	181	NRRMHNSFTADNRRATILIGRNIGDEYFKVEDTYFADLLILATGSGVYGEVSHDFDRYWA	240			
QY	241	SHSAHNATRIIRSGNIGGGLQALGYNDETSRHALLRIRFETVEQSPLYOKIOTGRIDMOSV	300			
Db	241	SHSAHNATRIIRSGNIGGGLQALGYNDETSRHALLRIRFETVEQSPLYOKIOTGRIDMOSV	300			
QY	301	QTRLISDDPAAGLDRKKPPLAGLQDALQPEKSVLYLSEFYFVPTKSGDIALAKLVQD	360			
Db	301	QTRLISDDPAAGLDRKKPPLAGLQDALQPEKSVLYLSEFYFVPTKSGDIALAKLVQD	360			

Dd	301	QTRLLSDDBAKGLDBDRKKRPPFAGRLQDALQOPEKSVYLVSEFYFPTTSGDIALAKLYQD	360
Qy	361	GIDVTVLNLSLOATDVAAVHSGVYKRYRRPRLKAGIKYLELOPNNAVPATKDKGLGSSVT	420
Dd	361	GIDVTVLNLSLOATDVAAVHSGVYKRYRRPRLKAGIKYLELOPNNAVPATKDKGLGSSVT	420
Qy	421	SIHAATFIVDGRKIFIGSFNLDPBSARLNTENGVIYESPKIAEQMERLADTSEYAYRV	480
Dd	421	SIHAATFIVDGRKIFIGSFNLDPBSARLNTENGVIYESPKIAEQMERLADTSEYAYRV	480
Qy	481	TLDRHNRLQWMDPATRKTTYPNDEPEAKLMKRIAAKLTLSPLESLL	525
Dd	481	TLDRHNRLQWMDPATRKTTYPNDEPEAKLMKRIAAKLTLSPLESLL	525

RESULT 2  
B81083  
cardiolipin synthetase family protein NMBA1434 [imported] - *Neisseria meningitidis* (strain B81083)  
C:Species: *Neisseria meningitidis*  
C:Date: 31-Mar-2000 #sequence:revision 31-Mar-2000 #text:change 19-Jan-2001  
C:Accession: B81083  
R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.; H. Olin, H.; Vamthepan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A:Authors: Gradi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Aftile: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
A:Reference number: AB1000; MUID:2015755; PMID:10710307  
A:Accession: B81083  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <TEF>  
A:Cross-references: GB:AF002494; GB:AF002098; NID:g7226684; PIDN:AA41795.1; PID:g72222A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMBA1434

Query March 45.18: Score 237; DB 2: Length 508;  
Best Local Similarity 100.08; Pred. No. 7.1e-233; Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 MKTRSLISLCLLSCSCSMLPPLERTESRHFNFSKPFVRLDNIQIRHPTNGLSDIY 77  
DB 1 MKTRSLISLCLLSCSCSMLPPLERTESRHFNFSKPFVRLDNIQIRHPTNGLSDIY 60

QY 78 LINDPHEFAFAAALIEAESHLDIQYIWRNDISGRLLFNLVYLAERGVRRLLDDN 137  
DB 61 LINDPHEFAFAAALIEAESHLDIQYIWRNDISGRLLFNLVYLAERGVRRLLDDN 120

QY 138 NTRGIDDLILLADSPNIEVRLFNPFVLRKRRALGYLDPFRLNRHMKSEFTADNRATI 197  
DB 121 NTRGIDDLILLADSPNIEVRLFNPFVLRKRRALGYLDPFRLNRHMKSEFTADNRATI 180

QY 198 LGRNIGDEYKRVGEDTVFADLDILATGSVYGEVSHDEDRYAWASHAHNAATRIIRSG 254  
DB 181 LGRNIGDEYKRVGEDTVFADLDILATGSVYGEVSHDEDRYAWASHAHNAATRIIRSG 237

RESULT 3  
AE0636  
conserved hypothetical protein STY1185 [imported] - *Salmonella enterica* subsp. *enterica*  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: this species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AE0636  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connelton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar  
S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AE0636

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-495 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD08272.1; PID:916503219; GSPDB:GN00176  
 A:Gene: STY1185

Query Match 2.9%; Score 15; DB 2; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 LDSPNIEVRLNPF 163  
 |||||  
 DB 116 LDSPNIEVRLNPF 130

## RESULT 4

probable synthase [imported] - Escherichia coli (strain O157:H7, substrain RMD 050952)  
 C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: H90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
 A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: H90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834847.1; PID:913360888; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RMD 050952

C:Genetics:

A:Gene: ECs1424

Query Match 2.7%; Score 14; DB 2; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDDNNT 139  
 |||||  
 DB 91 RGVRRLLDDNNT 104

## RESULT 5

probable synthase ymdc [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
 C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C:Accession: D85666

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 11ler, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dinalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: D85666

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <STO>

A:Cross-references: GB:AE005174; NID:912514578; PIDN:AMC55792.1; GSPDB:GN00145; UWGP:216

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: ymdc

Query Match 2.7%; Score 14; DB 2; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDDNNT 139  
 |||||  
 DB 91 RGVRRLLDDNNT 104

## RESULT 6

ymdc protein - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli

C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: C64847

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
 Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: C64847

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-493 <BLAT>

A:Cross-references: GB:AE000206; GB:U00096; NID:91787282; PIDN:AMC74130.1; PID:917872

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: ymdc

Query Match 2.7%; Score 14; DB 2; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDDNNT 139  
 |||||  
 DB 91 RGVRRLLDDNNT 104

## RESULT 7

hypothetical 55.9K protein in csgc-mdog intergenic region [imported] - Agrobacterium

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: G97536

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,  
 Science 294, 2323-2328, 2001

A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A87359; PMID:11743194

A:Accession: G97536

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-518 <KUR>

A:Cross-references: GB:AE007869; PIDN:AK67408.1; PID:915156719; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_3007

A:Map position: circular chromosome

Query Match 2.1%; Score 11; DB 2; Length 518;  
 Best Local Similarity 100.0%; Pred. No. 0.016;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDD 136  
 |||||  
 DB 115 RGVRRLLDD 125

## RESULT 8

phospholipase D family protein [imported] - Agrobacterium tumefaciens (strain C58, Du

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002

C:Accession: AB2777

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AB2777  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-518 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AL42632.1; PID:917740062; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atul630  
A:Map position: circular chromosome

Query Match 2.1%; Score 11; DB 2; Length 518;  
Best Local Similarity 100.0%; Pred. No. 0.016;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDD 136  
|||||  
Db 115 RGVRRLLDD 125

## RESULT 9

135403  
Probable polyprenyl synthetase - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 21-Jan-2000  
C:Accession: 135403  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: 221577  
A:Accession: 135403  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-378 <OLIT>

A:Cross-references: EMBL:AL049485; PIDN:CAB39696.1; GSPDB:GN00070; SCOEDB:SC6A5.12  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC6A5.12  
C:Superfamily: prenyl transferase A

Query Match 1.7%; Score 9; DB 2; Length 378;  
Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 FAARALIE 94  
|||||  
Db 322 FAARALIE 330

## RESULT 10

AC1951  
hypothetical protein alr1158 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AC1951  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yanada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC1951  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-161 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA073115.1; PID:917130504; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr1158

C:Superfamily: Synechococcus hypothetical protein 4 (uncl 5' region)

Query Match 1.5%; Score 8; DB 2; Length 161;  
Best Local Similarity 100.0%; Pred. No. 6.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 YITDFPRL 180  
|||||  
Db 141 YITDFPRL 148

## RESULT 11

EB6950  
conserved hypothetical protein ML0333 [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001  
C:Accession: EB6950  
R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; R.; Davies, R.M.; Devlin, K.; Duttoy, S.; Feltsell, T.; Fraser, A.; Hamlin, N.; Holtoeam, M.A.; Rutherford, K.M.  
Nature 409, 1007-1011, 2001  
A:Authors: Rutherford, K.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; A:Title: Massive gene decay in the leprosy bacillus.  
A:Reference number: AB6909; MUID:21128732; PMID:11234002  
A:Accession: EB6950  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-252 <STO>

A:Cross-references: GB:AL450380; NID:913092638; PIDN:CAC29841.1; GSPDB:GN00147  
C:Genetics:  
A:Gene: ML0333  
C:Superfamily: Bacillus subtilis lactam utilization protein ycsf

Query Match 1.5%; Score 8; DB 2; Length 252;  
Best Local Similarity 100.0%; Pred. No. 9.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 LAERGV 129  
|||||  
Db 52 LAERGV 59

## RESULT 12

S49376  
hypothetical protein 3 - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa  
C:Date: 16-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 08-Oct-1999  
C:Accession: S49376  
R:Group, B.; Hunderer, C.; Jahn, D.  
submitted to the EMBL Data Library, October 1994  
A:Description: The Escherichia coli hemN gene encoding oxygen-independent coproporphyrin synthetase.  
A:Reference number: S49376  
A:Accession: S49376  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-576 <TOU>  
A:Cross-references: EMBL:X82071; NID:9557256; PIDN:CA57572.1; PID:9557259

Query Match 1.5%; Score 8; DB 2; Length 576;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 PLKAGIK 396  
|||||  
Db 241 PLKAGIK 248

## RESULT 13

D83063  
hypothetical protein PA4667 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83063  
R:Stover, C.K.; Pham, X.Q.; Errin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adam, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.  
A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: DB3063  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-590 <STO>  
 A:Cross-references: GB:AE004680; GB:AE004091; NID:g9950912; PIDN:AG08054.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA4667

Query Match  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 PLKAGIK 396  
 |||||  
 DB 256 PLKAGIK 263

RESULT 14  
 B82724  
 cardiolipin synthase XF1087 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: B82724  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen  
 Nature 406, 151-157, 2000  
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A>Note: for a complete list of authors see reference number A59328 below  
 A:Accession: B82724  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-652 <SIM>  
 A:Cross-references: GB:AE003445; GB:AE003849; NID:g9106036; PIDN:AAF3897.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carriero, D.M.; Carreir, H  
 as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GeneBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; Laigt  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiti, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Saneli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF1087

Query Match  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLL 133  
 |||||  
 DB 126 RGVRRLL 133

RESULT 15  
 G95929  
 Probable methylcrotonyl-CoA carboxylase (EC 6.4.1.4) [imported] - Sinorhizobium meliloti  
 C:Species: Sinorhizobium meliloti  
 C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001  
 C:Accession: G95929  
 R:Flan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A>Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; MUID:21396508; PMID:11481431  
 A:Accession: G95929  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-662 <KUR>  
 A:Cross-references: GB:AU591985; PIDN:CAC49103.1; PID:g15140588; GSPDB:GN00167  
 A:Experimental source: strain 1021, megaplasmid pSymB  
 R:Gallbert, F.; Flan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barley-Hubi  
 pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: mcbB; SMB21124  
 A:Genome: plasmid  
 C:Superfamily: propionyl-CoA carboxylase alpha chain; biotin carboxylase homology; 11  
 C:Keywords: ligase

Query Match  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 324 GRLOALK 331  
 |||||  
 DB 405 GRLOALK 412

RESULT 16  
 D38164  
 cobD protein - Pseudomonas sp.  
 C:Species: Pseudomonas sp.  
 C>Date: 31-Jul-1992 #sequence\_revision 31-Jul-1992 #text\_change 08-Oct-1999  
 C:Accession: D38164  
 R:Crouzet, J.; Levy-Schli, S.; Cameron, B.; Cauchols, L.; Rigault, S.; Rouyez, M.C.;  
 J. Bacteriol. 173, 6074-6087, 1991  
 A>Title: Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair Pseudomonas  
 adenosyltransferase, cobD, and bifunctional cobinamide kinase-cobin  
 A:Reference number: A38164; MUID:92011366; PMID:1655697  
 A:Accession: D38164  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1275 <CRO>  
 A:Cross-references: GB:M62866; NID:g151150; PIDN:AAA25780.1; PID:g151154  
 C:Superfamily: Rhodobacter capsulatus magnesium-protocorphylin O-methyltransferase

Query Match  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 388 KPLKAGI 395  
 |||||  
 DB 170 KPLKAGI 177

RESULT 17  
 A96616  
 unknown protein F16M22.5 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: A96616  
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
 nansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzia  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MUID:21016719; PMID:11130712  
 A:Accession: A96616

A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2599 <STO>  
 A:Cross-References: GB:AE005173; NID:g11055836; PIDN:AA628304.1; GSPDB:GN00141  
 C:Genetics:  
 A:Gene: F16M22.5  
 A:Map position: 1

Query Match 1.5%; Score 8; DB 2; Length 2599;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SDDPAKGL 313  
 DB 1516 SDDPAKGL 1523

RESULT 18  
 T31307

type I fatty acid synthase homolog - *Cryptosporidium parvum*.

C:Species: *Cryptosporidium parvum*

C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Nov-2000

C:Accession: T31307

R:Zhu, G.; Marchewka, M.J.; Woods, K.M.; Upton, S.J.; Keithly, J.S.

submitted to the EMBL Data Library, August 1998

A:Description: Characterization of a type I FAS gene in the parasitic protozoan *Cryptosporidium*

A:Reference number: Z20993

A:Accession: T31307

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-8243 <ZHU>

A:Cross-References: EMBL:AF082993; NID:g4092068; PID:g4092069; PIDN:AAC99407.1

C:Genetics:

A>Note: FAS1

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology; acetate-CoA ligase

C:Keywords: carrier protein

F:100-714/Domain: acetate-CoA ligase homology <ACLI>

F:791-861/Domain: acyl carrier protein homology <ACP1>

F:906-1308/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS1>

F:2924-2992/Domain: acyl carrier protein homology <ACP2>

F:3062-3463/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS2>

F:5157-5227/Domain: acyl carrier protein homology <ACP3>

F:5274-5689/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>

Query Match 1.5%; Score 8; DB 2; Length 8243;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 GLTGSSVT 420  
 DB 1045 GLTGSSVT 1052

RESULT 19  
 A70693

hypothetical protein RV2830C - *Mycobacterium tuberculosis* (strain H37RV)

C:Species: *Mycobacterium tuberculosis*

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: A70693

R:Coile, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70693

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-71 <COL>

A:Cross-References: GB:Z8131; GB:AL133456; NID:g3261650; PIDN:CAB03656.1; PID:e276785;

A:Experimental source: strain H37RV  
 C:Genetics:

A:Gene: RV2830C

Query Match 1.3%; Score 7; DB 2; Length 71;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 AKILSL 519  
 DB 8 AKILSL 14

RESULT 20  
 S29563

endothelin 2 precursor - dog (fragment)

C:Species: *Canis lupus familiaris* (dog)

C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 16-Jul-1999

C:Accession: S29563; S04854

R:Itch, Y.

submitted to the EMBL Data Library, December 1990

A:Description: Nucleotide sequence of a canine endothelin-2 gene segment encoding the

A:Reference number: S29563

A:Accession: S29563

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-77 <ITO>

A:Cross-References: EMBL:X57038; NID:8896; PIDN:CAA40354.1; PID:9897

R:Itch, Y.; Kimura, C.; Onda, H.; Fujino, M.

Nucleic Acids Res. 17, 5389, 1989

A:Title: Canine endothelin-2: cDNA sequence for the mature peptide.

A:Reference number: S04854; MUID:89345104; PMID:2668883

A:Accession: S04854

A:Molecule type: mRNA

A:Residues: 22-77 <IT2>

A:Cross-References: EMBL:X15067

C:Superfamily: endothelin

F:30-56/Product: big endothelin 2 #status predicted <MA2>

F:30-51/Product: endothelin 2 #status predicted <MA2>

Query Match 1.3%; Score 7; DB 2; Length 77;

Best Local Similarity 100.0%; Pred. No. 37;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CSCSSWL 38  
 DB 52 CSCSSWL 58

RESULT 21  
 S09751

hypothetical protein TR12 - human cytomegalovirus (strain AD169)

C:Species: human cytomegalovirus, human herpesvirus 5

A>Note: host Homo sapiens (man)

C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 08-Oct-1999

C:Accession: S09751

R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horenell, T.

M.; Barrell, B.G.

Curr. Top. Microbiol. Immunol. 154, 125-169, 1990

A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09751

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-115 <CHD>

A:Cross-References: EMBL:X17403; NID:g59591; PIDN:CAA35450.1; PID:e298598; PID:g17808

A>Note: this sequence was submitted to the EMBL Data Library, December 1989

A>Note: this reading frame extends between two stop codons and does not begin with a

Query Match 1.3%; Score 7; DB 2; Length 115;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 LITCLLC 32  
 LITCLLC 32



Db 73 LILCLLC 79

RESULT 22

E83964

hypothetical protein BH2517 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: E83964

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E83964

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-117 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g510174886; PIDN:BAB06236.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2517

Query Match 1.3%; Score 7; DB 2; Length 117;

Best Local Similarity 100.0%; Pred. No. 53;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 LISLCL 29

|||||

Db 81 LISLCL 87

RESULT 23

G72716

hypothetical protein APE1150 - *Aeropyrum pernix* (strain K1)

C:Species: *Aeropyrum pernix*

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999

C:Accession: G72716

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*

A:Reference number: A72450; MUID:99310339; PMID:10382966

A:Accession: G72716

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-125 <KAW>

A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA80135.1; PID:d1043921; PID:g510

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1150

Query Match 1.3%; Score 7; DB 2; Length 125;

Best Local Similarity 100.0%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 GLOALGX 265

|||||

Db 31 GLOALGX 37

RESULT 24

T31900

hypothetical protein T05H4.8 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T31900

R:Blanchard, M.

submitted to the EMBL Data Library, July 1997

A:Description: The sequence of *C. elegans* cosmid T05H4.

A:Reference number: 221097

A:Accession: T31900

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-127 <BLA>

A:Cross-references: EMBL:AF016452; PIDN:AB66014.1; GSPDB:GN00023; CESP:T05H4.8

A:Experimental source: strain Bristol N2; clone T05H4

C:Genetics:

A:Gene: CESP:T05H4.8

A:Map position: 5

A:Introns: 30/3; 42/3; 62/2; 103/1

Query Match 1.3%; Score 7; DB 2; Length 127;

Best Local Similarity 100.0%; Pred. No. 57;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 TESRHF 51

|||||

Db 121 TESRHF 127

RESULT 25

F64712

penicillin-binding protein 4 - *Helicobacter pylori* (strain 26695)

C:Species: *Helicobacter pylori*

C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 28-Jul-2000

C:Accession: F64712

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, F.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliq, H.G.; Glodek, A.; McKe

son, J.D.; Kelley, J.M.; Cotton, M.D.; Feldman, J.M.; Fujiki, C.; Bowman, C.; Watthey,

Nature 388, 539-547, 1997

A:Authors: Mallin, E.; Hayes, W.S.; Borodovsky, M.; Karp, P.D.; Smith, H.O.; Fraser,

A:Title: The complete genome sequence of the gastric pathogen *Helicobacter pylori*.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: F64712

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-136 <TOM>

A:Cross-references: GB:AE000652; GB:AE000511; NID:g2314720; PIDN:AAD08589.1; PID:g231

C:Superfamily: Campylobacter jejuni hypothetical protein Cj1086c

Query Match 1.3%; Score 7; DB 2; Length 136;

Best Local Similarity 100.0%; Pred. No. 60;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 TGSVGE 230

|||||

Db 58 TGSVGE 64

RESULT 26

D71806

hypothetical protein jhp1457 - *Helicobacter pylori* (strain J99)

C:Species: *Helicobacter pylori*

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 28-Jul-2000

C:Accession: D71806

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dohg, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557; PMID:9923682

A:Accession: D71806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-136 <ARN>

A:Cross-references: GB:AE001567; GB:AE001439; NID:g4156065; PIDN:AAD07031.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp1457

C:Superfamily: Campylobacter jejuni hypothetical protein Cj1086c

Query Match 1.3%; Score 7; DB 2; Length 136;

Best Local Similarity 100.0%; Pred. No. 60;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 TGSVGE 230

DB 58 TGSVGE 64

# RESULT 27

568817 legumin beta chain - Japanese cedar (fragment)

C:Species: Cryptomeria japonica (Japanese cedar)

C>Date: 04-Dec-1997 #sequence\_revision 04-Dec-1997 #text\_change 07-May-1999

C:Accession: S68817

R:Haeger, K.P.; Mueller, B.; Wind, C.; Erbach, S.; Fischer, H.

FEBS Lett. 387, 94-98, 1996

A:Title: Evolution of legumin genes: loss of an ancestral intron at the beginning of ang

A:Reference number: S68817; MUID:96234018; PMID:8654576

A:Accession: S68817

A:Molecule type: DNA

A:Residues: 1-148 <HAE>

A:Experimental source: leaves

C:Function:

A:Description: seed storage protein

C:Superfamily: glycinin

C:Keywords: seed; storage protein

Query Match

Best Local Similarity 100.0%; Pred. No. 65;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# RESULT 28

568818 legumin beta chain - giant arborvitae (fragment)

C:Species: Thuja plicata (giant arborvitae)

C>Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 07-May-1999

C:Accession: S68818

R:Haeger, K.P.; Mueller, B.; Wind, C.; Erbach, S.; Fischer, H.

FEBS Lett. 387, 94-98, 1996

A:Title: Evolution of legumin genes: loss of an ancestral intron at the beginning of ang

A:Reference number: S68812; MUID:96234018; PMID:8654576

A:Accession: S68818

A:Molecule type: DNA

A:Residues: 1-148 <HAE>

A:Experimental source: leaves

C:Function:

A:Description: seed storage protein

C:Superfamily: glycinin

C:Keywords: seed; storage protein

Query Match

Best Local Similarity 100.0%; Pred. No. 65;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# RESULT 29

E97502 riboflavin synthase (A1132928) [imported] - Agrobacterium tumefaciens (strain C58, Cerec

C:Species: Agrobacterium tumefaciens

C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: E97502

R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lapps, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: E97502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-148 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86974.1; PID:q15156212; GSPDB:GN00169

C:Genetics:

A:Gene: AGR\_C\_2165

A:Map position: circular chromosome

C:Superfamily: riboflavin synthase beta chain

Query Match

Best Local Similarity 100.0%; Pred. No. 65;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# RESULT 30

A12720 riboflavin synthase beta chain [imported] - Agrobacterium tumefaciens (strain C58, Du

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002

C:Accession: A12720

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: A12720

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-148 <KUR>

A:Cross-references: GB:AE008688; PIDN:AA142183.1; PID:q17739573; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: ribH

A:Map position: circular chromosome

C:Superfamily: riboflavin synthase beta chain

Query Match

Best Local Similarity 100.0%; Pred. No. 65;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

# RESULT 31

AE3400 riboflavin synthase (EC 2.5.1.9) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 19-Apr-2002

C:Accession: AE3400

R:DeLVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Muier, C.; Ios, T.; Ivanov

; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688

A:Accession: AE3400

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-157 <KUR>

A:Cross-references: GB:AE008917; PIDN:AA152368.1; PID:q17983166; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME1187

A:Map position: I

C:Superfamily: riboflavin synthase beta chain

C:Keywords: transferase

Query Match 1.3%; Score 7; DB 2; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 86 FAARAL 92  
 1111111  
 Db 140 FAARAL 146

## RESULT 32

C84290  
 hypothetical protein Vng1355h [imported] - Halobacterium sp. NRC-1.  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: C84290  
 R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabid Jung, K.H.; Alam, M.; Freltas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; MUID:20504483; PMID:11016950  
 A:Accession: C84290  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-159 <STO>  
 A:Cross-references: GB:AE004437; NID:g10580867; PIDN:AG19687.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: VNG1355H  
 C:Superfamily: Streptomyces coelicolor probable integral membrane protein SCE126.11

Query Match 1.3%; Score 7; DB 2; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 272 HALIRYR 278  
 1111111  
 Db 58 HALIRYR 64

## RESULT 33

T48767  
 hypothetical protein 13E11.150 [imported] - Neurospora crassa  
 C:Species: Neurospora crassa  
 C>Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 19-May-2000  
 C:Accession: T48767  
 R:Schulte, U.; Align, V.; Hohnsels, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, April 2000  
 A:Reference number: Z24541  
 A:Accession: T48767  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-161 <SCH>  
 A:Cross-references: EMBL:AL353820; GSPDB:GN00112; NCSP:13E11.150  
 A:Experimental source: cosmid contig 13E11; strain 74  
 C:Genetics:  
 A:Gene: NCSP:13E11.150  
 A:Map position: 2  
 A:introns: 87/2  
 C:Superfamily: Neurospora crassa hypothetical protein 13E11.150

Query Match 1.3%; Score 7; DB 2; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 LAKLYD 360  
 1111111  
 Db 144 LAKLYD 150

## RESULT 34

F69992

thioredoxin peroxidase (EC 1.11.1.-) ytg1 - Bacillus subtilis  
 N:Alternate names: scavengase; thiol peroxidase p20  
 C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: F69992

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; A.; Ehrlich, S.D.; Emmerman, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gal lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Huilo, J. Koelter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidis, A.; Lardin A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mai y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scantl A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S. akouch, M.; Tamakoshi, A.; Tanaka, T.; Terpsita, P.; Tognoni, A.; Tosato, V.; Uchlyr T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: F69992  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-167 <KUN>  
 A:Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CB14927.1; PID:g2635. A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ytg1  
 C:Superfamily: thioredoxin peroxidase  
 C:Keywords: oxidoreductase; redox-active disulfide  
 F:60-94/Disulfide bonds: redox-active #status predicted

Query Match 1.3%; Score 7; DB 2; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 365 TVLTNSL 371  
 1111111  
 Db 28 TVLTNSL 34

## RESULT 35

AF0803  
 probable lipoprotein SRY2607 [imported] - Salmonella enterica subsp. enterica serovar A:Species: Salmonella enterica subsp. enterica serovar Typhi  
 A:Note: this species has also been called Salmonella typhi  
 C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
 C:Accession: AF0803  
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair , S.; Moule, S.; O'Garra, P.  
 Nature 413, 848-852, 2001  
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s A:Reference number: AB0502; PMID:11677608  
 A:Accession: AF0803  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-172 <PAR>  
 A:Cross-references: GB:AL513382; PIDN:CAD07608.1; PID:g16503600; GSPDB:GN00176  
 C:Genetics:  
 A:Gene: SRY2607

Query Match 1.3%; Score 7; DB 2; Length 172;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 29 LLLCSCS 35  
 1111111  
 Db 11 LLLCSCS 17

```

RESULT 36
A39070
A:Molecule type: precursor - human
N:Contains: big endothelin 2; endothelin 2; endothelin homolog
C:Species: Homo sapiens (man)
C>Date: 20-Mar-1992 #sequence_revision 26-May-1994 #text_change 18-Jun-1999
R:Accession: A39070; S12912; A61252; B32111
R:Bioch, K.D.; Hong, C.C.; Eddy, R.L.; Shows, T.B.; Quertermous, T.
Genomics 10, 236-242, 1991
A:Title: cDNA cloning and chromosomal assignment of the endothelin 2 gene: vasoactive in
A:Reference number: A39070; MUID:91257834; PMID:1840558
A:Accession: A39070
A:Molecule type: mRNA
A:Residues: 1-178 <BLO>
A:Cross-references: GB:M65200
R:Ohkubo, S.; Ogi, K.; Hosoya, M.; Matsumoto, H.; Suzuki, N.; Kimura, C.; Onda, H.; Fujii
FEBS Lett. 274, 136-140, 1990
A:Title: Specific expression of human endothelin-2 (ET-2) gene in a renal adenocarcinoma
A:Reference number: S12912; MUID:91071415; PMID:1701397
A:Accession: S12912
A:Molecule type: mRNA
A:Residues: 1-178 <ONR>
A:Cross-references: GB:X55177; NID:931258; PIDN:CAA38962.1; PID:931259
R:Onda, H.; Ohkubo, S.; Kosaka, T.; Yasuhara, T.; Ogi, K.; Hosoya, M.; Matsumoto, H.; Su
J. Cardiovasc. Pharmacol. 17(Suppl. 7), S39-S43, 1991
A:Title: Expression of endothelin-2 (ET-2) gene in a human renal adenocarcinoma cell lin
A:Reference number: A61252; MUID:92219747; PMID:1725387
A:Accession: A61252
A:Molecule type: mRNA
A:Residues: 1-178 <ONR>
R:Inoue, A.; Yanagisawa, M.; Kimura, S.; Kasuya, Y.; Miyachi, T.; Goto, K.; Masaki, T.
Proc. Natl. Acad. Sci. U.S.A. 86, 2863-2867, 1989
A:Title: The human endothelin family: three structurally and pharmacologically distinct
A:Reference number: A32111; MUID:89202426; PMID:2649896
A:Accession: B32111
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 23-73 <INO>
A:Cross-references: GB:M25550; GB:J04522; NID:9556204; PIDN:AAA52340.1; PID:9556205
C:Genetics:
A:Gene: GDB:EDN2
A:Cross-references: GDB:125241; OMIM:131241
A:Map position: 1p34-1p34
C:Superfamily: endothelin
C:Keywords: vasoconstrictor
F:1-24/Domain: signal sequence #status predicted <SIG>
F:49-85/Product: big endothelin 2 #status predicted <BMAT>
F:96-110/Product: endothelin 2 #status predicted <MAT>
F:49-63,51-59,96-110,98-106/Disulfide bonds: #status predicted
Query Match 1.3%; Score 7; DB 1; Length 178;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 32 CSCSML 38
DB 49 CSCSML 55

```

```

A:Molecule type: DNA
A:Residues: 1-178 <SEN>
A:Cross-references: EMBL:U60315; NID:91491943; PIDN:AA55243.1; PID:91492058
C:Genetics:
A:Note: MC115L
C:Superfamily: vaccinia virus 20K virion protein
Query Match 1.3%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 228 VGEVSHD 234
DB 109 VGEVSHD 115
RESULT 38
T32796
A:Molecule type: protein F14D2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
R:Du, Z.; Lee, T.T.
Submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F14D2.
A:Reference number: T32796
A:Accession: T32796
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-178 <DUZ>
A:Cross-references: EMBL:AF040643; PIDN:AB94958.1; GSPDB:GN00020; CESP:F14D2.2
A:Experimental source: strain Bristol N2; clone F14D2
C:Genetics:
A:Gene: CESP:F14D2.2
A:Map position: 2
A:Introns: 94/1
Query Match 1.3%; Score 7; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 166 RKRALG 172
DB 75 RKRALG 81
RESULT 39
B64167
cytochrome c biogenesis protein Cycx homolog ccmg precursor [similarity] - Haemophilu
N:Alternate names: disulfide interchange protein dsdb; thiol-disulfide interchange pr
C:Species: Haemophilus influenzae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
R:Flisbachman, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Goeyne, J.D.; Scott, J.; Shiley, R.; Liu, L.T.; Glodex, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64167
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-181 <TIGR>
A:Cross-references: GB:U022789; GB:U42023; NID:91574642; PIDN:AA022752.1; PID:91574650
C:Genetics:
A:Gene: dsdb
C:Superfamily: cytochrome c biogenesis protein Cycx
C:Keywords: redox-active disulfide
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-181/Product: cytochrome c biogenesis protein cycx homolog ccmg #status predicted
F:78-81/Disulfide bonds: redox-active #status predicted

```

Query Match 1.3%; Score 7; DB 1; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 426 TFIYDCK 432  
 |||||  
 DB 144 TFIYDCK 150

## RESULT 40

AB3472  
 Hypothetical protein BMEI1760 [Imported] - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 C:Accession: AB3472  
 R:DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Muier, C.; Los, T.; Ivanova,  
 .; Mazur, M.; Goltzman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: AB3472  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-188 <NR>  
 A:Cross-References: GB:AE008917; PIDN:AAL52941.1; PID:g17983791; GSPDB:GN00190  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BMEI1760  
 A:Map position: I

Query Match 1.3%; Score 7; DB 2; Length 188;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 315 RDRKKP 321  
 |||||  
 DB 21 RDRKKP 27

Search completed: May 12, 2003, 10:09:16  
 Job time : 72 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:01:11 ; Search time 14 Seconds  
(without alignments)  
1555.362 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 525  
Sequence: 1 MRANKTOAMPSETISLMKT.....KLMKRIAAKILSLPIESLL 525

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	14	2.7	493	1 YMDC_ECOLI	P75919 escherichia
2	8	1.5	272	1 HIS6_METKA	O81708 methanopyru
3	8	1.5	590	1 YAG7_PSEAE	P42810 pseudomonas
4	8	1.5	1275	1 COBN_PSEDE	P29929 pseudomonas
5	7	1.3	56	1 ET2_CANFA	P12064 canis famill
6	7	1.3	115	1 IR02_HCMVA	P16802 human cytom
7	7	1.3	130	1 RS6E_METKA	O81706 methanopyru
8	7	1.3	153	1 RIB1_RHIME	O92400 rhizobium m
9	7	1.3	166	1 TPX_BACSU	P80864 bacillus su
10	7	1.3	178	1 ET2_HUMAN	P20800 homo sapien
11	7	1.3	181	1 DSBE_HAEIN	P45038 haemophilus
12	7	1.3	192	1 LPCA_HELPJ	O92421 helicobacte
13	7	1.3	211	1 YCSE_BACSU	P42863 bacillus su
14	7	1.3	221	1 PSPA_ECOLI	P23853 escherichia
15	7	1.3	238	1 RL2_SUISO	O94045 sulfolobus
16	7	1.3	247	1 RS6_APICA	O94045 aplysia cal
17	7	1.3	253	1 RL2_SCHPO	P08093 schizosacch
18	7	1.3	253	1 RL2_YEAST	P05736 saccharomyc
19	7	1.3	254	1 HA2R_HUMAN	P01903 homo sapien
20	7	1.3	263	1 Y683_HALNI	O94112 halobacteri
21	7	1.3	290	1 TSNX_HUMAN	O94998 homo sapien
22	7	1.3	293	1 BHCL_RHOGO	P47331 rhodococcus
23	7	1.3	293	1 Y766_TREPA	O83747 treponema p
24	7	1.3	313	1 META_VIBCH	O94705 vibrio chol
25	7	1.3	322	1 Y486_METJA	O57910 methanococ
26	7	1.3	354	1 E2B_XYLFA	O94995 xyliella fas
27	7	1.3	371	1 SRD1_CAEEL	O19992 caenorhabd
28	7	1.3	374	1 IDRA_CAEEL	O93714 caenorhabd
29	7	1.3	375	1 GBB3_TOBAC	O40507 nicotiana t
30	7	1.3	375	1 VANC_ARATH	O94587 arabidopsis
31	7	1.3	377	1 GBB1_TOBAC	P93397 nicotiana t
32	7	1.3	377	1 GBB2_TOBAC	P93398 nicotiana t
33	7	1.3	377	1 GBB_NICPL	P93339 nicotiana p

34	7	1.3	377	1 GBB_SOUTU	P93563 solanum tub
35	7	1.3	380	1 GBB_MATZE	P49178 zea mays (m
36	7	1.3	380	1 GBB_ORYSA	O40687 oryza sativ
37	7	1.3	380	1 SCH4_STRHA	O05361 streptomyce
38	7	1.3	385	1 ISG4_STRCO	O94799 streptomyce
39	7	1.3	394	1 NDRI_HUMAN	O92597 homo sapien
40	7	1.3	394	1 NDRI_MOUSE	O62433 mus musculu
41	7	1.3	396	1 DDL_TREPA	O83676 treponema p
42	7	1.3	412	1 THIM_RAT	P13437 rattus norv
43	7	1.3	422	1 RBAL_CAEEL	P90917 caenorhabd
44	7	1.3	430	1 AROA_STRPY	P23491 mesocricetu
45	7	1.3	436	1 RP54_BACSU	O99283 streptococ
46	7	1.3	454	1 MSRE_RABIT	P24219 bacillus su
47	7	1.3	457	1 ARLY_ECO57	O05585 erythrolagus
48	7	1.3	457	1 ARLY_ECO57	O84730 escherichia
49	7	1.3	457	1 ARLY_ECO57	P11447 escherichia
50	7	1.3	458	1 TRME_STRPY	O99200 streptococ
51	7	1.3	467	1 GAC3_HUMAN	O99928 homo sapien
52	7	1.3	467	1 GAC3_MOUSE	P27681 mus musculu
53	7	1.3	467	1 GAC3_RAT	P28473 rattus norv
54	7	1.3	500	1 GABT_RAT	P50554 rattus norv
55	7	1.3	502	1 Y190_HELPJ	O92402 helicobacte
56	7	1.3	502	1 Y190_HELPJ	P56117 helicobacte
57	7	1.3	516	1 TRPE_MYCTU	O06127 mycobacteri
58	7	1.3	528	1 SERA_MYCLE	O33116 mycobacteri
59	7	1.3	528	1 SERA_MYCTU	O53243 mycobacteri
60	7	1.3	528	1 V5BK_BSMV	P04867 barley stri
61	7	1.3	529	1 TRPE_MYCLE	O94705 escherichia
62	7	1.3	542	1 CAB2_METKA	O94996 thermoplas
63	7	1.3	556	1 ADEC_METJA	O81708 methanopyru
64	7	1.3	566	1 SYFB_BORBU	O58854 methanococ
65	7	1.3	573	1 DPOL_MOUSE	P94283 borrelia bu
66	7	1.3	575	1 DPOL_HUMAN	O94928 mus musculu
67	7	1.3	589	1 PPCK_THEVO	O94925 homo sapien
68	7	1.3	601	1 TORX_HUMAN	P58306 thermoplas
69	7	1.3	634	1 SEJB_MOOTH	O95996 homo sapien
70	7	1.3	742	1 ZW10_ARATH	O46435 mootella th
71	7	1.3	803	1 NASD_BACSU	O48626 arabidopsis
72	7	1.3	815	1 TTKA_DROME	P42282 drosophila
73	7	1.3	830	1 GYRA_BUCAI	P57277 buchnera ap
74	7	1.3	873	1 FAS2_DROME	P34482 drosophila
75	7	1.3	1537	1 DNMI_CHICK	O92072 gallus gall
76	7	1.3	1616	1 DNMI_HUMAN	P26358 homo sapien
77	7	1.1	42	1 YPAL_ECOLI	P03394 escherichia
78	6	1.1	48	1 Y18_BPT7	P03727 bacterioph
79	6	1.1	68	1 VHTU_LAMB	P51256 porphyra pu
80	6	1.1	70	1 RRI8_PORPU	P51256 porphyra pu
81	6	1.1	72	1 CHH_PROBO	P55445 procambartus
82	6	1.1	73	1 RRI8_GUTTH	O78488 guillardia
83	6	1.1	74	1 MIFH_PROBO	O10987 procambartus
84	6	1.1	74	1 NIFH_NOSSN	P52336 nostoc sp.
85	6	1.1	74	1 Y299_TREPA	O83321 treponema p
86	6	1.1	78	1 IFIC_MARPO	P12123 marchantia
87	6	1.1	83	1 PRPR_RAT	P81278 rattus norv
88	6	1.1	88	1 Y874_CAUCR	P58091 caulobacter
89	6	1.1	92	1 RRI8_EPIVI	P30061 epifagus vi
90	6	1.1	94	1 CH10_CLOPE	O26322 clostridium
91	6	1.1	97	1 SY07_MOUSE	P33666 mus musculu
92	6	1.1	97	1 SY08_MOUSE	O94221 mus musculu
93	6	1.1	97	1 Y404_ANASP	O84793 anabaena sp
94	6	1.1	98	1 PRPR_BOVIN	P81264 bos taurus
95	6	1.1	98	1 SRGI_MOUSE	O88745 mus musculu
96	6	1.1	98	1 SRGI_RAT	O92046 rattus norv
97	6	1.1	98	1 SY13_HUMAN	O99616 homo sapien
98	6	1.1	98	1 YX79_SYNY3	P73634 synechocyst
99	6	1.1	99	1 MCPA_BOVIN	P28291 bos taurus
100	6	1.1	99	1 SY02_HUMAN	P13500 homo sapien
101	6	1.1	99	1 SY02_HUMAN	O94994 macaca fasc
102	6	1.1	99	1 SY02_PIG	P42831 sus scrofa
103	6	1.1	99	1 SY07_HUMAN	P80098 homo sapien
104	6	1.1	99	1 SY08_HUMAN	P80075 homo sapien
105	6	1.1	99	1 SY08_PIG	P49673 sus scrofa
106	6	1.1	100	1 RRI8_PINTH	P52763 pinus thunb





253	6	1.1	206	1	SODM_MYCLE	P13367 mycobacteri	326	6	1.1	265	1	RL7A_CHICK	P32429 gallus gall
254	6	1.1	206	1	SODM_MYCLP	086165 mycobacteri	327	6	1.1	265	1	RL7A_FUGRU	057592 fuigu rubrip
255	6	1.1	207	1	ALAG_MOUSE	060590 mus musculu	328	6	1.1	265	1	RL7A_HUMAN	P15158 homo sapien
256	6	1.1	207	1	COAE_PSEPU	P36644 pseudomonas	329	6	1.1	265	1	RL7A_ICTPU	090942 ictalutis p
257	6	1.1	208	1	RL13_SCHPO	074175 schizosacch	330	6	1.1	265	1	RL7A_MOUSE	P12970 mus musculu
258	6	1.1	209	1	Y034_NPVOP	005126 orgyia pseu	331	6	1.1	266	1	UPK_PYRAE	082940 pyrobaculum
259	6	1.1	210	1	YGBL_HAEIN	057199 haemophilus	332	6	1.1	267	1	ZEAE_MAIZE	P04668 zea mays (m
260	6	1.1	212	1	TAG2_MOUSE	09wva4 mus musculu	333	6	1.1	268	1	RP30_HUMAN	P78346 homo sapien
261	6	1.1	213	1	RIB7_SULSO	P95872 sulfolobus	334	6	1.1	268	1	Y002_YEAST	P36025 saccharomyc
262	6	1.1	216	1	SOMA_MOUSE	P06880 mus musculu	335	6	1.1	269	1	AGAR_ECOLI	P42902 escherichia
263	6	1.1	216	1	SOMA_RAT	P01244 rattus norv	336	6	1.1	269	1	REFX_MOUSE	092205 mus musculu
264	6	1.1	216	1	Y458_METJA	057500 methanococc	337	6	1.1	269	1	VC03_SPEKA	P32229 swinepox vi
265	6	1.1	217	1	FK21_NEUCR	060046 neuropept	338	6	1.1	270	1	CATA_RHOOP	P95607 rhodococcus
266	6	1.1	217	1	GI27_FASHE	P31670 fasciola he	339	6	1.1	270	1	NORO_PARDE	051654 paracoccus
267	6	1.1	217	1	TER4_ECOLI	P09164 escherichia	340	6	1.1	274	1	Y772_TREPA	083751 treponema p
268	6	1.1	218	1	CPJ3_CAPAE	P79152 capra aegag	341	6	1.1	276	1	TCPN_VIBCH	P29492 vibrio chol
269	6	1.1	218	1	TER8_PASPI	P51562 pasteurella	342	6	1.1	277	1	HEKM_BUCAI	P57269 buchnera ap
270	6	1.1	221	1	PUR0_DEIRA	094783 delnococtus	343	6	1.1	277	1	Y087_THEMA	09wx42 thetmotoga
271	6	1.1	222	1	RL4_CHLMU	09pj15 chlamydia m	344	6	1.1	278	1	Y541_CHLPN	092811 chlamydia p
272	6	1.1	222	1	RL4_CHLTR	084532 chlamydia t	345	6	1.1	278	1	YF17_METJA	058912 methanococc
273	6	1.1	225	1	VE26_NPVAC	P12827 autographa	346	6	1.1	279	1	TONB_HAEDU	051810 haemophilus
274	6	1.1	225	1	DCTR_BACSU	P96602 bacillus su	347	6	1.1	281	1	MODB_HAEIN	057278 haemophilus
275	6	1.1	227	1	COX2_CERGA	P98020 cercocobus	348	6	1.1	282	1	TAUD_ECOLI	P37610 escherichia
276	6	1.1	227	1	COX2_MANLE	P98037 manorillius	349	6	1.1	282	1	YFHH_ECOLI	P37767 escherichia
277	6	1.1	227	1	UNG_MYCLE	Q9cbs3 mycobacteri	350	6	1.1	284	1	THTM_PSEAE	Q91452 pseudomonas
278	6	1.1	229	1	ISPD_NEIMA	Q9jtm3 neisseria m	351	6	1.1	284	1	TRUA_STRCO	P37767 streptomyce
279	6	1.1	229	1	ISPD_NEIMB	Q9jym4 neisseria m	352	6	1.1	287	1	RPOF_STRCO	P37971 streptomyce
280	6	1.1	229	1	MODB_HAEIN	P45322 haemophilus	353	6	1.1	287	1	UL24_ILTVT	P23966 infectious
281	6	1.1	229	1	YTG7_CAEEL	011083 caenorhabdi	354	6	1.1	287	1	YCXE_BACSU	P40420 bacillus su
282	6	1.1	231	1	UPPS_AOUAE	067291 aquifex aeo	355	6	1.1	288	1	ACCD_PORPU	P51198 porphyra pu
283	6	1.1	233	1	MTRE_HUMAN	Q15012 homo sapien	356	6	1.1	288	1	HMXX_CHICK	P50223 gallus gall
284	6	1.1	233	1	MTRE_MOUSE	060861 mus musculu	357	6	1.1	288	1	SP4G_BACSU	P26937 bacillus su
285	6	1.1	234	1	EFPU_PANMO	P50380 pandorina m	358	6	1.1	288	1	Y504_CHLTR	084512 chlamydia t
286	6	1.1	236	1	ALDC_IACLA	P95676 lactococcus	359	6	1.1	290	1	COBD_ARCFU	028923 archaeglob
287	6	1.1	236	1	ALDC_IACLC	P71890 lactococcus	360	6	1.1	290	1	VP19_HCMVA	P15673 human cytom
288	6	1.1	236	1	DSBC_ECOLI	P21892 escherichia	361	6	1.1	291	1	BACH_NATPH	P15647 natronomona
289	6	1.1	236	1	KDGR_ERWCH	Q93493 erwania chr	362	6	1.1	291	1	TB02_YEAST	P38282 saccharomyc
290	6	1.1	236	1	YV99_CIOAB	P33664 clostridium	363	6	1.1	292	1	VBL1_BGMV	Q91179 salmonella
291	6	1.1	236	1	YVDE_IACLO	Q9c600 iactococcus	364	6	1.1	293	1	VBL1_SALTY	091179 bean golden
292	6	1.1	238	1	IAHI_YEAST	P41734 saccharomyc	365	6	1.1	293	1	VBL1_TGMV	P06001 nostoc comm
293	6	1.1	238	1	OMPA_CITFR	P24016 citrobacter	366	6	1.1	294	1	NIEH_ANASL	P33178 anabaena sp
294	6	1.1	239	1	RS6B_SCHPO	P05752 schizosacch	367	6	1.1	295	1	NIEH_NOSS6	051296 nostoc sp.
295	6	1.1	239	1	RS6B_SCHPO	Q9c027 schizosacch	368	6	1.1	295	1	NIEH_NOSS6	P00457 anabaena sp
296	6	1.1	242	1	YD26_HAEIN	P44162 haemophilus	369	6	1.1	295	1	NIEH_NOSS6	Q43866 anabaena va
297	6	1.1	244	1	RS6_BRAFL	001727 salmonello	370	6	1.1	295	1	NIEH_NOSS6	Q47917 mastigoclad
298	6	1.1	245	1	CBIM_SALTY	005594 salmonello	371	6	1.1	296	1	KHRI_YEAST	P22313 saccharomyc
299	6	1.1	245	1	IF6_DROME	P56388 drosophila	372	6	1.1	296	1	NIEH_PLEBO	000240 plectonema
300	6	1.1	248	1	DAT_YEAST	P13483 saccharomyc	373	6	1.1	296	1	NIEH_PLEBO	055028 synechococc
301	6	1.1	248	1	RS6_DROME	P29327 drosophila	374	6	1.1	296	1	NIEH_PLEBO	Q44484 anabaena va
302	6	1.1	249	1	RS6_ARATH	P51430 arabidopsis	375	6	1.1	297	1	GPDA_CAMJE	Q94439 campylobact
303	6	1.1	249	1	RS6_CHICK	P47838 gallus gall	376	6	1.1	297	1	NIEH_NOSS6	P26250 nostoc comm
304	6	1.1	249	1	RS6_HUMAN	P10660 homo sapien	377	6	1.1	297	1	NIEH_NOSS6	Q30577 anabaena sp
305	6	1.1	249	1	RS6_HUMAN	Q90978 ictalutis p	378	6	1.1	299	1	ICIA_AERSA	P70773 aeromonas s
306	6	1.1	249	1	RS6_ONCMY	Q99172 oncorhynchus	379	6	1.1	300	1	BLAB_PROVU	P52664 proteus vul
307	6	1.1	249	1	RS6_XENLA	P39017 xenopus lae	380	6	1.1	300	1	RNH3_CHLTR	Q84011 chlamydia t
308	6	1.1	249	1	T2FB_HUMAN	P13984 homo sapien	381	6	1.1	302	1	CYSD_ECOLI	P21156 escherichia
309	6	1.1	251	1	T2FB_RAT	Q01750 rattus norv	382	6	1.1	302	1	RIMK_HAEIN	P45241 haemophilus
310	6	1.1	251	1	RS6_ASPOF	Q9m398 asparagus o	383	6	1.1	303	1	HYVE_ECOLI	P73335 escherichia
311	6	1.1	253	1	YIEE_ECOLI	P31464 escherichia	384	6	1.1	303	1	PYRD_PYRHO	059185 pyrococcus
312	6	1.1	254	1	CTRL_HALRO	P35003 halloctis ru	385	6	1.1	305	1	ACCD_STNPF	054776 synechococc
313	6	1.1	254	1	YA95_MCPN	P75597 mycoplasma	386	6	1.1	305	1	P2A_HELAN	P48579 hellanthus
314	6	1.1	255	1	RL8A_YEAST	P17076 saccharomyc	387	6	1.1	306	1	YE16_YEAST	Q44011 streptomyce
315	6	1.1	255	1	RL8B_YEAST	P29455 saccharomyc	388	6	1.1	307	1	MEFE_STRLI	054235 streptomyce
316	6	1.1	255	1	YARF_ECOLI	Q47147 escherichia	389	6	1.1	307	1	P2A3_ORISA	Q94977 oryza sativ
317	6	1.1	257	1	TRPC_AOUAE	067657 aquifex aeo	390	6	1.1	307	1	P2A3_ARATH	Q04951 arabidopsis
318	6	1.1	257	1	Y453_AOUAE	066760 aquifex aeo	391	6	1.1	308	1	DDL_AGRIS	Q8udn3 agrobacteri
319	6	1.1	258	1	HIS6_ECOLI	P10373 escherichia	392	6	1.1	308	1	MP5A_LOLPR	Q40240 lolium pere
320	6	1.1	258	1	HIS6_SALTY	P10373 salmonella	393	6	1.1	309	1	NARS_MOUSE	P79332 mus musculu
321	6	1.1	258	1	LLDR_ECOLI	P33233 escherichia	394	6	1.1	309	1	PNAD_MOUSE	Q64311 mus musculu
322	6	1.1	258	1	LLDR_ECOLI	Q12189 saccharomyc	395	6	1.1	309	1	PNAD_MOUSE	Q28955 sus scrofa
323	6	1.1	261	1	TFXG_RHILT	P42729 rhizobium 1	396	6	1.1	312	1	K1PF_ECOLI	P23539 escherichia
324	6	1.1	261	1	Y133_MYCGE	P47379 mycoplasma	397	6	1.1	312	1	PYRB_AERPE	Q9yba4 aeropyrum p
325	6	1.1	264	1	U2AG_DROME	Q94535 drosophila	398	6	1.1	313	1	YJ19_AGRIS	P58770 agrobacteri



545	6	1.1	377	1	GBB_ARATH	P49177	arabidopsis	618	6	1.1	425	1	YC18_METUA	Q58615	methanococc
546	6	1.1	377	1	HIS8_MYCLE	Q9x7b8	mycobacteri	619	6	1.1	425	1	YF31_METUA	Q58922	methanococc
547	6	1.1	377	1	YBHS_ECOLI	P75775	eschlerichia	620	6	1.1	428	1	HEMY_HAEIN	P44727	haemophilus
548	6	1.1	377	1	YD80_AGRF5	O8uf17	agrobacteri	621	6	1.1	428	1	PYRC_BACSU	P25995	bacillus su
549	6	1.1	379	1	YD8B_ECOLI	P11027	eschlerichia	622	6	1.1	428	1	Y486_MYCLE	P34138	mycobacteri
550	6	1.1	379	1	VP26_YEAST	P40335	saccharomyc	623	6	1.1	430	1	CAE1_DROME	Q24572	drosophila
551	6	1.1	380	1	H181_MYCTU	O06591	mycobacteri	624	6	1.1	431	1	YD1B_CHLFR	O48604	chlamydia t
552	6	1.1	380	1	MODC_AZOVI	P37732	azotobacter	625	6	1.1	431	1	TOLE_SCHPO	014021	schizosacch
553	6	1.1	384	1	CYNX_ECOLI	P17583	eschlerichia	626	6	1.1	435	1	DHOM_METGL	P37143	methylobact
554	6	1.1	384	1	ISG1_STRCO	O9x7w2	streptomyce	627	6	1.1	435	1	GA8L_YEAST	P04387	saccharomyc
555	6	1.1	385	1	BIOF_MYCLE	P45487	mycobacteri	628	6	1.1	436	1	GD6F_BOVIN	P55106	bos taurus
556	6	1.1	385	1	PAR4_HUMAN	Q96r10	homo sapien	629	6	1.1	437	1	IFCY_ECOLI	P76503	eschlerichia
557	6	1.1	385	1	RBA2_CAEEL	P90916	caenorhabdi	630	6	1.1	437	1	FEZG_METUA	Q58657	methanococc
558	6	1.1	385	1	YL90_MYCTU	Q10383	mycobacteri	631	6	1.1	438	1	SRMB_LACHE	P94870	lactobacill
559	6	1.1	386	1	BIOF_MYCTU	O06621	mycobacteri	632	6	1.1	439	1	SRMB_HAEIN	P44701	haemophilus
560	6	1.1	386	1	YAMW_SCHPO	Q10191	schizosacch	633	6	1.1	439	1	YDWM_ECOLI	P76130	eschlerichia
561	6	1.1	388	1	CARA_ANASP	O8yxq7	anaeana sp	634	6	1.1	442	1	TRME_MYCTU	Q98rj5	mycoplasma
562	6	1.1	388	1	VENV_MCV1	P26579	moliscum c	635	6	1.1	443	1	TDCC_ECOLI	P11867	eschlerichia
563	6	1.1	394	1	LPXB_SYNY3	Q57310	synchocyst	636	6	1.1	444	1	SRMB_ECOLI	P21507	eschlerichia
564	6	1.1	395	1	PAR4_RAT	Q920e0	rattus norv	637	6	1.1	445	1	FLDL_VIBPA	Q03475	vibrio para
565	6	1.1	396	1	AMOA_AERYH	P23300	aeromonas h	638	6	1.1	447	1	EF1A_TOBAC	P43643	nicotiana t
566	6	1.1	396	1	518L_HUMAN	O75177	homo sapien	639	6	1.1	448	1	DCPD_RHILE	P10046	rhizobium l
567	6	1.1	396	1	YANG_RHISN	P55579	rhizobium s	640	6	1.1	448	1	EF1A_LYCES	P17786	lycopersico
568	6	1.1	398	1	POR_DAUCA	Q9sdt1	daucus caro	641	6	1.1	449	1	TRN4_ECOLI	Q60216	eschlerichia
569	6	1.1	399	1	GCAM_MOUSE	P01865	mus musculu	642	6	1.1	449	1	EF11_DAUCA	P29521	daucus caro
570	6	1.1	401	1	PORC_ARATH	P21218	arabidopsis	643	6	1.1	449	1	EF1A_ARATH	P13905	arabidopsis
571	6	1.1	401	1	PORC_ARATH	O48741	arabidopsis	644	6	1.1	450	1	CSK_CHICK	P41239	gallus gall
572	6	1.1	401	1	TAMI_POVNA	P03070	hamster pol	645	6	1.1	450	1	CY1_PARDE	P13627	paracoccus
573	6	1.1	402	1	GP24_HUMAN	Q99705	homo sapien	646	6	1.1	451	1	GNTF_ZYMMO	Q92670	zymomonas m
574	6	1.1	402	1	VN34_ROTFC	P27586	porcine tot	647	6	1.1	451	1	NFS1_MOUSE	Q92670	mus musculu
575	6	1.1	403	1	T230_CAEEL	O09474	caenorhabdi	648	6	1.1	451	1	NFS1_RAT	Q99393	rattus norv
576	6	1.1	405	1	C130_MYCTU	Q11062	mycobacteri	649	6	1.1	451	1	YJBE_YEAST	P47051	saccharomyc
577	6	1.1	405	1	DCDA_HELPY	O9zmes	helicobacte	650	6	1.1	452	1	RN18_HUMAN	Q09s80	homo sapien
578	6	1.1	405	1	DCDA_HELPY	P56129	helicobacte	651	6	1.1	453	1	SHR1_APICA	O16550	aplysia cal
579	6	1.1	406	1	GCBM_MOUSE	P01867	mus musculu	652	6	1.1	453	1	Y259_MYCRN	P75419	mycoplasma
580	6	1.1	406	1	TPSP_HUMAN	P05154	homo sapien	653	6	1.1	454	1	TKS3_HUMAN	P57727	homo sapien
581	6	1.1	408	1	MTV1_VIBS3	Q03055	vibrio sp.	654	6	1.1	455	1	PGKC_CRIFA	P08967	critihidia f
582	6	1.1	408	1	PKR_THEVO	Q97bc6	thermoplasm	655	6	1.1	455	1	TRME_LACLA	Q9cch8	lactococcus
583	6	1.1	408	1	YEIJ_SCHPO	Q13883	schizosacch	656	6	1.1	456	1	FXD1_MOUSE	Q61343	mus musculu
584	6	1.1	409	1	TERA_PSESP	P33009	pseudomonas	657	6	1.1	456	1	SAT_ARCFU	Q28606	archaeoglob
585	6	1.1	410	1	ODPX_YEAST	P16451	saccharomyc	658	6	1.1	456	1	Y883_HAEIN	P44971	haemophilus
586	6	1.1	411	1	AIAT_RAT	P17475	rattus norv	659	6	1.1	457	1	DEBA_ECOLI	P21693	eschlerichia
587	6	1.1	411	1	ACCD_CHLVU	P56293	chlorella v	660	6	1.1	457	1	NFS1_HUMAN	Q9y697	homo sapien
588	6	1.1	411	1	APGM_PYRFU	P58814	pyrococcus	661	6	1.1	458	1	HRA2_HUMAN	Q43644	homo sapien
589	6	1.1	412	1	AIAT_MUSCR	P26595	mus caroli	662	6	1.1	459	1	COAT_FLDV	P22172	fish lympho
590	6	1.1	412	1	APGM_PYRHO	O57742	pyrococcus	663	6	1.1	459	1	TRME_ANASP	O8y9n1	anaeana sp
591	6	1.1	412	1	EX7L_ANASP	O8y7w4	anaeana sp	664	6	1.1	460	1	NIFN_RHILC	Q98ap3	rhizobium l
592	6	1.1	412	1	PHOA_PENCH	P37274	penicillium	665	6	1.1	460	1	YAGG_ECOLI	P75683	eschlerichia
593	6	1.1	414	1	RCA_ANASP	P58555	anaeana sp	666	6	1.1	461	1	DLDH_CHLPN	Q92773	chlamydia p
594	6	1.1	414	1	THIK_YARLI	Q05493	yarowia li	667	6	1.1	461	1	EL3B_WHEAT	P52409	triticum ae
595	6	1.1	414	1	YH87_YEAST	Q04835	saccharomyc	668	6	1.1	461	1	RBB4_MOUSE	O60972	mus musculu
596	6	1.1	415	1	LEU2_SULFO	Q974r0	sulfolobus	669	6	1.1	461	1	YAO1_HUMAN	O95521	homo sapien
597	6	1.1	415	1	RCA_ANASC	O06721	anaeana sp	670	6	1.1	462	1	TRME_STRAC	O9r197	streptococc
598	6	1.1	417	1	PGKB_CRIFA	P08966	critihidia f	671	6	1.1	463	1	DESM_CHICK	P02542	gallus gall
599	6	1.1	418	1	RHLB_HAEIN	P44922	haemophilus	672	6	1.1	463	1	Y102_MYCTU	O53951	mycobacteri
600	6	1.1	418	1	SSXT_HUMAN	Q15532	homo sapien	673	6	1.1	464	1	RCCL_CANAL	P52499	candida alb
601	6	1.1	418	1	SSXT_MOUSE	Q62280	mus musculu	674	6	1.1	465	1	DLDH_CHLMU	O9p013	chlamydia m
602	6	1.1	419	1	KDAP_MOUSE	O09043	mus musculu	675	6	1.1	465	1	DLDH_CHLFR	O84561	chlamydia t
603	6	1.1	420	1	MUCB_ECOLI	P07375	eschlerichia	676	6	1.1	465	1	GACL_MOUSE	O940y8	mus musculu
604	6	1.1	420	1	RHLB_ECOLI	P24229	eschlerichia	677	6	1.1	465	1	GACL_RAT	P23574	rattus norv
605	6	1.1	420	1	RHLB_SALTY	P40863	salmoneila	678	6	1.1	466	1	FAY_HUMAN	P08709	homo sapien
606	6	1.1	420	1	YA39_PYRHO	O58758	pyrococcus	679	6	1.1	466	1	GAC2_RAT	P18108	rattus norv
607	6	1.1	421	1	5H72_APICA	O16591	aplysia cal	680	6	1.1	467	1	DADR_HUMAN	P21917	homo sapien
608	6	1.1	421	1	MUCB_SALTY	P14303	salmoneila	681	6	1.1	467	1	E2BD_SCHPO	O09924	schizosacch
609	6	1.1	424	1	MS11_ARATH	O22467	arabidopsis	682	6	1.1	467	1	GAC2_HUMAN	P18507	homo sapien
610	6	1.1	424	1	MS11_LYCES	O22466	lycopersico	683	6	1.1	467	1	MBB2_METUA	P58871	methanosarc
611	6	1.1	424	1	SP60_DICDI	P15270	dactylostele	684	6	1.1	467	1	MBB3_METUA	P58872	methanosarc
612	6	1.1	424	1	VK04_VACCC	P20537	vaccinia vi	685	6	1.1	467	1	SMA2_HUMAN	Q15796	homo sapien
613	6	1.1	424	1	VK04_VACCV	P18377	vaccinia vi	686	6	1.1	467	1	SMA2_MOUSE	O62433	rattus norv
614	6	1.1	424	1	ZP3_MOUSE	P10761	mus musculu	687	6	1.1	467	1	SMA2_RAT	O70436	rattus norv
615	6	1.1	425	1	RBB4_HUMAN	O09028	h chromatin	688	6	1.1	467	1	SYC_ARCFU	O29936	archaeoglob
616	6	1.1	425	1	RBB7_HUMAN	Q16576	homo sapien	689	6	1.1	468	1	MBB2_METAC	O85772	methanosarc
617	6	1.1	425	1	RBB7_MOUSE	Q60973	mus musculu	690	6	1.1	468	1	VI2_HPV51	P26539	human papill

691	6	1.1	469	1	CYTA_TRYEO	P26338	trypanosoma	764	6	1.1	520	1	YD99_LISMO	Q91738	listeria mo
692	6	1.1	471	1	RTN2_MOUSE	O70622	mus musculus	765	6	1.1	520	1	YMDA_BACSU	O31774	bacillus su
693	6	1.1	472	1	VL2_HPV34	P6758	human papill	766	6	1.1	521	1	DB45_DROME	O07886	drosophila
694	6	1.1	473	1	GAB3_HUMAN	P28472	homo sapien	767	6	1.1	521	1	P2BA_BOVIN	P48452	bos taurus
695	6	1.1	473	1	GAB3_MOUSE	P15433	mus musculus	768	6	1.1	521	1	P2BA_MOUSE	P284_MOUSE	mus musculus
696	6	1.1	473	1	YV33_MYCTU	O06250	mycobacteri	769	6	1.1	521	1	P2BA_MOUSE	P284_MOUSE	mus musculus
697	6	1.1	474	1	FLP_ZYGBA	P13769	zygosacchar	770	6	1.1	521	1	SCOT_CAEL	Q09450	caenorhabdi
698	6	1.1	474	1	GAC2_CHICK	P21548	gallus galli	771	6	1.1	521	1	YN78_BACHD	Q10934	caenorhabdi
699	6	1.1	474	1	GAC2_MOUSE	P22123	mus musculus	772	6	1.1	521	1	YT25_CAEL	Q10934	caenorhabdi
700	6	1.1	474	1	YAO5_HUMAN	O60811	homo sapien	773	6	1.1	522	1	ACCD_SPIOL	Q16222	h udp-n-ace
701	6	1.1	475	1	NIFB_AMAZ	P23300	bos taurus	774	6	1.1	522	1	UAP1_HUMAN	O18222	h udp-n-ace
702	6	1.1	475	1	NIFB_AMAZ	P23300	bos taurus	775	6	1.1	523	1	ASNS_YEAST	P20447	saccharomyc
703	6	1.1	475	1	NIFB_AMAZ	P23300	bos taurus	776	6	1.1	524	1	ASNS_SANAU	O24338	sandersonia
704	6	1.1	476	1	CUS_CLOPE	P20627	anabaena sp	777	6	1.1	524	1	EXP9_STRPN	O94066	arabidopsis
705	6	1.1	476	1	COX1_PLABE	O99252	plasmidium	778	6	1.1	524	1	EXP9_STRPN	P35599	streptococc
706	6	1.1	478	1	TIG_AOUAE	O67358	aquifex aeo	779	6	1.1	524	1	P2BB_HUMAN	P14299	homo sapien
707	6	1.1	479	1	DBPA_BACSU	P42205	bacillus su	780	6	1.1	524	1	TY3H_CAEL	P69086	caenorhabdi
708	6	1.1	480	1	Y486_MYCTU	O11152	mycobacteri	781	6	1.1	525	1	BRB2_HUMAN	O94770	homo sapien
709	6	1.1	481	1	PHR_HALHA	P20377	halobacteri	782	6	1.1	525	1	P2BB_RAT	P20651	rattus norv
710	6	1.1	482	1	CAB1_METJA	O58773	methanococc	783	6	1.1	526	1	BUTY_HUMAN	Q13410	homo sapien
711	6	1.1	482	1	DHPC_PSESP	P19059	pseudomonas	784	6	1.1	527	1	RHGB_ASAPC	O00019	aspergillus
712	6	1.1	486	1	VPEB_ARATH	Q33044	arabidopsis	785	6	1.1	532	1	PMGI_SYNY3	P74507	synechocyst
713	6	1.1	486	1	XYLG_PSEPU	P23105	pseudomonas	786	6	1.1	534	1	TCPG_YEAST	P39077	saccharomyc
714	6	1.1	488	1	ACCD_ARATH	P56765	arabidopsis	787	6	1.1	538	1	TCPG_YEAST	O94066	arabidopsis
715	6	1.1	488	1	CATA_LISIN	O926x0	listeria in	788	6	1.1	538	1	NADB_PSEAE	O51363	pseudomonas
716	6	1.1	488	1	CATA_LISIN	O926x0	listeria in	789	6	1.1	538	1	THIP_HAELN	P44985	haemophilus
717	6	1.1	488	1	CATA_LISIN	O926x0	listeria in	790	6	1.1	540	1	CH60_MYCPA	P47828	candida alb
718	6	1.1	488	1	DHAL_PSESP	P33008	pseudomonas	791	6	1.1	540	1	TCPO_CANAL	P47828	candida alb
719	6	1.1	488	1	YAL5_HAELN	P71364	haemophilus	792	6	1.1	540	1	YFEO_YEAST	P4562	saccharomyc
720	6	1.1	489	1	LEF9_NPVOP	O10319	oryzias pseu	793	6	1.1	540	1	YK26_YEAST	P36112	saccharomyc
721	6	1.1	490	1	FLP_ZYXRO	P13785	zygosacchar	794	6	1.1	543	1	RRP3_YEAST	P36112	saccharomyc
722	6	1.1	492	1	DPR2_HUMAN	O94014	homo sapien	795	6	1.1	544	1	C821_PEA	Q43068	plum saliv
723	6	1.1	493	1	ACCD_EPIYI	P30064	epifagus vi	796	6	1.1	544	1	CH60_AERSA	O63039	aeromonas s
724	6	1.1	493	1	AMPA_AOUAE	O67868	aquifex aeo	797	6	1.1	544	1	TCPE_THERAC	P33044	thermoplasma
725	6	1.1	496	1	POLS_IBDYE	P29802	avian infec	798	6	1.1	544	1	TCPG_DROME	P46368	drosophila
726	6	1.1	497	1	ACCD_CUSRE	P31562	cuscuta ref	799	6	1.1	544	1	TCPG_HUMAN	P46368	homo sapien
727	6	1.1	499	1	CIMA_METKA	O8941	methanopyru	800	6	1.1	545	1	TCPG_MOUSE	O72298	homo sapien
728	6	1.1	499	1	GAK_MOUSE	O9944	mus musculus	801	6	1.1	545	1	TCPG_MOUSE	P80318	mus musculus
729	6	1.1	499	1	PITB_ECOLI	P43676	escherichia	802	6	1.1	546	1	DBP2_YEAST	P24783	saccharomyc
730	6	1.1	501	1	CP36_RABIT	P11707	oryctolagus	803	6	1.1	547	1	TCPE_TERYP	O15891	tetrahymena
731	6	1.1	502	1	C72G_ARATH	O91tm6	arabidopsis	804	6	1.1	547	1	TCPG_XENLA	P50143	xenopus lae
732	6	1.1	502	1	C72H_ARATH	O91tm6	arabidopsis	805	6	1.1	547	1	YOUN_BACSU	P54551	bacillus su
733	6	1.1	503	1	CUS_BACFI	O66043	bacillus fi	806	6	1.1	548	1	AAK1_RAT	Q13131	homo sapien
734	6	1.1	503	1	CUS_BACHD	O94824	bacillus ha	807	6	1.1	550	1	AAK1_HUMAN	O13131	homo sapien
735	6	1.1	503	1	CP39_RAT	P51538	rattus norv	808	6	1.1	550	1	STE_THERAC	O91m5	thermoplasma
736	6	1.1	503	1	CP39_CANFA	P24463	canis famli	809	6	1.1	553	1	MCRX_METTH	P21110	methanobact
737	6	1.1	503	1	CP3D_MOUSE	O64464	mus musculus	810	6	1.1	557	1	CNE6_HUMAN	O95741	homo sapien
738	6	1.1	503	1	YD39_YEAST	O03034	saccharomyc	811	6	1.1	557	1	CNE6_MOUSE	O95741	homo sapien
739	6	1.1	504	1	YM68_YEAST	O04991	saccharomyc	812	6	1.1	558	1	GGPI_DROTA	P52030	drosophila
740	6	1.1	505	1	PCLI_HUMAN	O9403	homo sapien	813	6	1.1	558	1	GGPI_DROTA	P52030	drosophila
741	6	1.1	506	1	SYG_DEIRA	O9485	deinococcus	814	6	1.1	558	1	YKXA_ASTLO	P52031	drosophila
742	6	1.1	507	1	CP3S_BOVIN	P79102	bos taurus	815	6	1.1	559	1	ECM1_MOUSE	P34782	atastia ion
743	6	1.1	507	1	HD42_CAEL	O09440	caenorhabdi	816	6	1.1	559	1	TCPG_OYXGR	O61508	mus musculus
744	6	1.1	509	1	CPV1_BRARE	O42145	brachydanio	817	6	1.1	559	1	TCPG_TERYP	O00782	oxytricha g
745	6	1.1	510	1	DHAF_VIBHA	O56694	vibrio harv	818	6	1.1	560	1	INRL_BOVIN	O04790	bos taurus
746	6	1.1	511	1	ACH2_RAT	P12389	rattus norv	819	6	1.1	560	1	INRL_SHEEP	O28589	ovis aries
747	6	1.1	512	1	ACCD_TOBAC	P12219	nicotiana t	820	6	1.1	561	1	DNLI_HAUNI	Q11033	haemobacteri
748	6	1.1	513	1	DMP1_HUMAN	Q13316	homo sapien	821	6	1.1	563	1	DEAD_MYCTU	O11033	haemobacteri
749	6	1.1	513	1	HEMO_CHICK	P18080	gallus galli	822	6	1.1	564	1	CAB1_METKA	O8470	methanopyru
750	6	1.1	514	1	PMGI_ECOLI	P37689	escherichia	823	6	1.1	568	1	HE64_TRYBB	O26696	trypanosoma
751	6	1.1	514	1	SYX_METMP	O30520	methanococc	824	6	1.1	568	1	MEND_PASMU	O94914	p menaquin
752	6	1.1	515	1	6GD4_DROME	O94914	drosophila	825	6	1.1	569	1	ILIR_HUMAN	P14778	homo sapien
753	6	1.1	515	1	P2BB_MOUSE	P48453	mus musculus	826	6	1.1	572	1	HOTU_STRCO	O94778	homo sapien
754	6	1.1	515	1	PTD9_FUSSO	Q16645	fusarium so	827	6	1.1	572	1	SYK_CAEL	O22099	caenorhabdi
755	6	1.1	515	1	STEF_YEAST	O67848	saccharomyc	828	6	1.1	573	1	CH60_CRICR	P18687	cricetulus
756	6	1.1	516	1	CBPX_ARATH	P32826	arabidopsis	829	6	1.1	573	1	CH60_DROME	O02649	drosophila
757	6	1.1	516	1	GLPD_MYCTU	O10502	mycobacteri	830	6	1.1	573	1	CH60_HUMAN	P19289	mus musculus
758	6	1.1	517	1	COX1_SULAC	P98004	sulfolobus	831	6	1.1	573	1	CH60_MOUSE	O49496	schizosacch
759	6	1.1	518	1	CPV2_CARAU	O7686	carassius a	832	6	1.1	573	1	CH60_MOUSE	O49496	schizosacch
760	6	1.1	518	1	LEI2_METJA	O58595	xylella fas	833	6	1.1	574	1	DEO2_SCHPO	P51520	simian retr
761	6	1.1	519	1	LEU1_XYLFA	O94993	xylella fas	834	6	1.1	575	1	ENV_SVR2R	O95817	homo sapien
762	6	1.1	520	1	LEU1_XANAC	P58900	xanthomonas	835	6	1.1	575	1	BAG3_HUMAN	O95817	homo sapien
763	6	1.1	520	1	LEU1_XANCP	P58901	xanthomonas	836	6	1.1	575	1	IUCA_ECOLI	O47316	escherichia



```

883 6 1.1 795 1 DD15_HUMAN 043143 homo sapien
884 6 1.1 798 1 PBPA_NEILA 087579 neisseria 1
885 6 1.1 798 1 PBPA_NEILA 005194 neisseria m
886 6 1.1 799 1 HIS2_SACBA 012670 saccharomyc
887 6 1.1 799 1 HIS2_YEAST P00815 saccharomyc
888 6 1.1 800 1 RR3_CHLEU P46307 chlamydomon
889 6 1.1 808 1 EXT2_HPANG 002381 hepatitis a
890 6 1.1 814 1 EXT2_CAEBL 001705 caenorhabdi
891 6 1.1 816 1 NEL_CHICK Q90827 gallus gall
892 6 1.1 819 1 SWEI_YEAST P32944 saccharomyc
893 6 1.1 827 1 PREL_RHOCA P23388 r multipilus
894 6 1.1 829 1 IF2_HAEIN P44323 haemophilus
895 6 1.1 829 1 VIR_A_AGR9 P10799 agrobacteri
896 6 1.1 829 1 VIR_A_AGR9 P07168 agrobacteri
897 6 1.1 831 1 IF2_RICPR Q92hf5 rickettsia
898 6 1.1 831 1 IF2_RICPR Q92hf5 rickettsia
899 6 1.1 833 1 IF2_PASMU P57873 pasteurella
1000 6 1.1 835 1 INVA_YEREN P19196 yersinia en

```

## ALIGNMENTS

## RESULT 1

YMDC\_ECOLI STANDARD: PRT: 493 AA.

```

ID YMDC_ECOLI STANDARD: PRT: 493 AA.
AC P/5919;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ymdc.
GN YMDC OR B1046.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RA MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY. CARDIOLIPIN
CC SYNTHASE SUBFAMILY. STRONG, TO H. PYLORI HP0190.
CC -1- SIMILARITY: CONTAINS 2 PLD PHOSPHODIESTERASE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; AE000206; AAC74130.1;
DR EMBL; D90741; BAA35836.1;

```

```

DR EMBL; D90742; BAA35844.1;
DR EcoGene; EG13875; ymdc.
DR InterPro; IPR001736; PLD.
DR Pfam; PF00614; PLDC; 2.
DR SMART; SM00155; PLDC; 2.
DR PROSITE; PS00035; PLD; 2.
KW Hypothetical protein; Transferase; Repeat; Complete proteome.
FT DOMAIN 145 172 PLD PHOSPHODIESTERASE 1.
FT DOMAIN 384 411 PLD PHOSPHODIESTERASE 2.
SQ SEQUENCE 493 AA; 55926 MW; 495604AD8DBD8492 CRC64;

```

Query Match 2.7%; Score 14; DB 1; Length 493;  
 Best Local Similarity 100.0%; Pred. No. 7; Gaps 0;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 126 RGVRRLLDDNNT 139
DB 91 RGVRRLLDDNNT 104

```

## RESULT 2

HIS6\_METKA STANDARD: PRT: 272 AA.

```

ID HIS6_METKA STANDARD: PRT: 272 AA.
AC Q8RYW8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisf (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisf) (IMGP synthase
DE subunit hisf) (IGP subunit hisf).
GN HISF OR MK0173.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_Taxid=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AV19 / DSM 6324 / JCM 9639;
RA MEDLINE-21927647; PubMed-11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbina-O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Nalele D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Maljkh A.G., Koonin E.V., Kozlovsk S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- FUNCTION: IGPs catalyzes the conversion of PREP and glutamine to
CC IGP, AICAR and glutamate. The hisf subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PREP using
CC the ammonia provided by the hish subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-(4S-phospho-1-deoxyribulose-1-
CC ylamino)methylidenesuccinyl-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2O).
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hish and hisf (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL; AE010316; AAM01390.1;
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT SITE 12 12 POTENTIAL.
FT ACT SITE 131 131 POTENTIAL.
SQ SEQUENCE 272 AA; 29848 MW; CA7F9018E3A1A0CC CRC64;

```

Query Match 1.5%; Score 8; DB 1; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 4.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 124 AERGVAVR 131  
 |||||  
 DB 263 AERGVAVR 270

## RESULT 3

YAG7\_PSEAE STANDARD; PRT; 590 AA.  
 ID YAG7\_PSEAE  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein PA4667.  
 GN PA4667.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE-95189718; PubMed-7883699;  
 RA Hungerer C., Troup B., Roemling U., Jahn D.;  
 RT "Regulation of the hema gene during 5-aminolevulinic acid formation  
 in Pseudomonas aeruginosa.";  
 RL J. Bacteriol. 177:1435-1443(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 15692 / PA01;  
 RX MEDLINE-20437337; PubMed-10984043;  
 RA Hickey M.J., Brinkman F.S.L., Mizoguchi S.D., Warren P.,  
 Garber R.L., Coltry L., Tolentino E., Westbrock-Wadman S., Yan Y.,  
 Barby L.L., Speller S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 Raeler J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----  
 CC EMBL: X82071; CAA57572.1; ALT\_INIT.  
 DR EMBL: AE004880; AAG08054.1; -.  
 DR InterPro: IPR001440; TPR: 5.  
 DR Pfam: PF00515; TPR: 5.  
 KW Hypothetical protein; Complete proteome.  
 FT CONFLICT 304 304 D -> DD (IN REF. 1).  
 SQ SEQUENCE 590 AA; 66290 MW; ABA9A48EB58497E CRC64;

Query Match 1.5%; Score 8; DB 1; Length 590;  
 Best Local Similarity 100.0%; Pred. No. 9.5;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 389 PLKAGIK 396  
 |||||  
 DB 256 PLKAGIK 263

## RESULT 4

COBN\_PSEDE STANDARD; PRT; 1275 AA.  
 ID COBN\_PSEDE  
 AC P29929;

DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE COBN protein.  
 GN COBN.  
 OS Pseudomonas denitrificans.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=43306;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-92011366; PubMed-1655697;  
 RA Crouzet J., Levy-Schli S., Cameron B., Cauchois L., Rigault S.,  
 Rouyez M.-C., Blanche F., Debussche L., Thibaut D.;  
 RT "Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair  
 Pseudomonas denitrificans DNA fragment containing five *cod* genes and  
 identification of structural genes encoding Cob(I)alamin  
 adenosyltransferase, cobyrilic acid synthase, and bifunctional  
 RT cobinamide kinase-cobinamide phosphate guanylyltransferase.";  
 RL J. Bacteriol. 173:6074-6087(1991).

CC -1- FUNCTION: PROPOSED TO PLAY A ROLE IN COBALT INSERTION REACTIONS.  
 CC -1- PATHWAY: cobalamin biosynthesis.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: M62866; AAA25780.1; -.  
 DR PIR: D38164; D38164.  
 DR InterPro: IPR003672; COBN/Mg\_chlase.  
 DR Pfam: PF002514; COBN-Mg\_chel\_1.  
 KW Cobalamin biosynthesis; Porphyrin biosynthesis.  
 SQ SEQUENCE 1275 AA; 138055 MW; 3FDF2858434E4C78 CRC64;

Query Match 1.5%; Score 8; DB 1; Length 1275;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 388 KPLKAGI 395  
 |||||  
 DB 170 KPLKAGI 177

RESULT 5  
 ET2\_CANFA STANDARD; PRT; 56 AA.  
 ID ET2\_CANFA  
 AC P12064;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Endothelin-2 precursor (ET-2) (Fragment).  
 GN EDN2.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE-89345104; PubMed-2668883;  
 RA Itoh Y., Kimura C., Onda H., Fujino M.;  
 RT "Canine endothelin-2: cDNA sequence for the mature peptide.";  
 RL Nucleic Acids Res. 17:5389-5389(1989).

CC -1- FUNCTION: ENDOTHELINS ARE ENDOTHELIUM-DERIVED VASOCONSTRICTOR  
 CC PEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO THE ENDOTHELIN/SARAFOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X57038; CAA40354.1; ALT\_SEQ.  
 DR PIR: S04854; S04854.  
 DR HSSP: P05305; 6CMH.  
 DR InterPro: IPR001928; Endothln\_tox.  
 DR Pfam: PF00322; endothelin; 1.  
 DR SMART: SM00272; EMD; 1.  
 DR PROSITE: PS00270; ENDOTHELIN; 1.  
 DR Cleavage on pair of basic residues; Vasoconstrictor; Multigene family.  
 FT NON\_TER 1 1  
 FT PEPTIDE 31 51 ENDOTHELIN-2.  
 FT DISULFID 31 45 BY SIMILARITY.  
 FT DISULFID 33 41 BY SIMILARITY.  
 SQ SEQUENCE 56 AA; 6112 MW; 1E63C56BC3DF04 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 32 CSCSSWL 38  
 DB 31 CSCSSWL 37

## RESULT 6

ID IR02\_HCMVA STANDARD; PRT; 115 AA.  
 AC P16802;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 01-FEB-1991 (Rel. 17, Last annotation update)  
 DE Hypothetical protein IRL2 (TRL2)  
 OS Human cytomegalovirus (strain AD169).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Cytomegalovirus.  
 OX NCBI\_TaxID=10360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90269039; PubMed=2161319;  
 RA Chee W.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
 RA Horsnell T., Hutchinson C.A. III, Kouzarides T., Martignetti J.A.,  
 RA Predde E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;  
 RT "Analysis of the protein-coding content of the sequence of human  
 RT cytomegalovirus strain AD169."  
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X17403; CAA35450.1; -  
 DR EMBL: X17403; CAA35308.1; -  
 DR PIR: S09751; S09751.  
 KW Hypothetical protein.  
 SQ SEQUENCE 115 AA; 12323 MW; FAE6EBD539B040C6 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 LUCILLC 32  
 DB 73 LUCILLC 79

## RESULT 7

ID R56E\_METKA STANDARD; PRT; 130 AA.  
 AC Q8YVE6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S6e.  
 GN RPS6E OR RPS6A OR MK1446.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AE010437; AAM02659.1; -  
 DR PROSITE: PS00578; RIBOSOMAL\_S6E; 1.  
 DR Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 130 AA; 14322 MW; 166ECAC9D304F624 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 VVRLL 134  
 DB 71 VVRLL 77

## RESULT 8

ID RIB1\_RHIME STANDARD; PRT; 153 AA.  
 AC Q920U0;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 6,7-dimethyl-8-ribityllumazine synthase 1 (EC 2.5.1.9) (DMRL synthase  
 DE 1) (Lumazine synthase 1) (Riboflavin synthase 1 beta chain).  
 GN RIB1 OR R01215 OR SMC01777.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,  
 RA Pohl T., Portetle D., Puehler A., Purnelle B., Rampeger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;



```

RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- FUNCTION: Riboflavin synthase is a bifunctional enzyme complex
CC catalyzing the formation of riboflavin from 5-amino-6-(1'-D)-
CC ribityl-amino-2,4(1H,3H)-pyrimidinedione and L-3,4-dihydroxy-2-
CC butanone-4-phosphate via 6,7-dimethyl-8-lumazine. The beta subunit
CC catalyzes the condensation of 5-amino-6-(1'-D)-ribityl-amino-
CC 2,4(1H,3H)-pyrimidinedione with L-3,4-dihydroxy-2-butanone-4-
CC phosphate yielding 6,7-dimethyl-8-lumazine (By similarity).
CC -1- CATALYTIC ACTIVITY: 2 6,7-dimethyl-8-(1'-D-ribityl)lumazine =
CC riboflavin + 4-(1'-D-ribitylamino)-5-amino-2,6-dihydroxypyrimidine.
CC -1- PATHWAY: Riboflavin biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE DMRL SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL591786; CAC45794.1;
DR InterPro: IPR002180; DMRL_synthase.
DR Pfam: PF00885; DMRL_synthase; 1.
DR ProDom: PD003664; DMRL_synthase; 1.
KW Riboflavin biosynthesis; Transferrase; Complete proteome.
SQ SEQUENCE 153 AA; 16193 MW; 8f50CA43DC559856 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 153;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 FARARAL 92
DB 134 FARARAL 140
|||||

RESULT 9
TPX_BACSU STANDARD; PRT; 166 AA.
AC P80864;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable thiol peroxidase (EC 1.11.1.-) (Superoxide-inducible protein
DE 8) (S018).
GN TPX.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98044033; PubMed=9387221;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Broillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,

```

```

RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA Joris B., Karmata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lavigne A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Plescan E., Puje P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serron S.J., Serron P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takegi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Tepstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vanlier F., Vassaretti A.,
RA Viati A., Wambutt R., Wedler E., Wedler H., Welzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yatsunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-31.
RC STRAIN=168 / IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis."
RL Electrophoresis 18:1451-1463(1997).
CC -1- FUNCTION: HAS ANTIOXIDANT ACTIVITY. COULD REMOVE PEROXIDES OR
CC H(2O(2)) (BY SIMILARITY).
CC -1- INDUCTION: BY SUPEROXIDE.
CC -1- SIMILARITY: BELONGS TO THE AHPc/TSA FAMILY. TPX SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF008220; AAC00316.1;
DR EMBL: 299119; CAB14927.1;
DR Subtilist; BG13855; tpx.
DR InterPro: IPR000866; AHPc-TSA.
DR Pfam: PF00578; AHPc-TSA; 1.
DR PROSITE: PS01265; TPX; 1.
KW Oxidoreductase; Peroxidase; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 166 AA; 18084 MW; 2BEPD4BFC6CA539 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 166;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 365 TVLTNSL 371
DB 27 TVLTNSL 33
|||||

RESULT 10
ET2_HUMAN STANDARD; PRT; 178 AA.
AC P20800;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Endothelin-2 precursor (ET-2).
GN EDN2.

```

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91071415; PubMed-1701397;  
 RA Okubo S., Ogi K., Hosoya M., Matsumoto H., Suzuki N., Kimura C.,  
 RT Onda H., Fujino M.,  
 RT "Specific expression of human endothelin-2 (ET-2) gene in a renal  
 RT adenocarcinoma cell line. Molecular cloning of cDNA encoding the  
 RT precursor of ET-2 and its characterization.",  
 RL FEBS Lett. 274:136-140(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91257834; PubMed-1840558;  
 RA Bloch R.D., Hong C.C., Eddy R.L. Jr., Shows T.B., Quettermous T.,  
 RT "cDNA cloning and chromosomal assignment of the endothelin 2 gene:  
 RT vasoactive intestinal contractor peptide is rat endothelin 2.",  
 RL Genomics 10:236-242(1991).  
 RN [3]  
 RP SEQUENCE OF 22-73 FROM N.A.  
 RX MEDLINE-89202426; PubMed-2649896;  
 RA Inoue A., Yanagisawa M., Kimura S., Kasuya Y., Miyachi T., Goto K.,  
 RA Masaki T.,  
 RT "The human endothelin family: three structurally and  
 RT pharmacologically distinct isopeptides predicted by three separate  
 RT genes.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2863-2867(1989).  
 CC -1- FUNCTION: ENDOTHELINS ARE ENDOTHELIUM-DERIVED VASOCONSTRICTOR  
 CC PEPTIDES.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE ENDOTHELIN/SARAFOTOXIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M65199; AAA52404.1; -  
 DR EMBL: X55177; CAA38962.1; -  
 DR EMBL: M25550; AAA52340.1; -  
 DR PIR: A39070; A39070.  
 DR PIR: B32111; B32111.  
 DR PIR: S12912; S12912.  
 DR HSP: P05305; 6CMH.  
 DR Genew; HGNC:3177; EDN2.  
 DR MIM: 131241; -  
 DR InterPro: IPR003641; Endothelin.  
 DR InterPro: IPR001928; Endothln\_tox.  
 DR Pfam: PF00322; endothelin\_1.  
 DR PRINTS: PR00365; Endothelin.  
 DR ProDom: PD005286; Endothelin; 1.  
 DR SMART: SM00272; END; 2.  
 DR PROSITE: PS00270; ENDOTHELIN; 2.  
 DR Cleavage on pair of basic residues; Vasoconstrictor; Multigene family;  
 KW Signal.  
 FT SIGNAL 1 24  
 FT PEPTIDE 49 69 POTENTIAL.  
 FT DOMAIN 96 111 ENDOTHELIN-2.  
 FT DISULFID 49 63 BY SIMILARITY.  
 FT DISULFID 51 59 BY SIMILARITY.  
 SQ SEQUENCE 178 AA: 19960 MW: 771182B6F075CFB4 CRC64;  
 Query Match 1.3%; Score 7; DB 1; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 32 CSCSSWL 38  
 |||||||

Db 49 CSCSSWL 55  
 RESULT 11  
 DSBE\_HAEIN  
 ID DSBE\_HAEIN STANDARD; PRT; 181 AA.  
 AC P45038;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Thiol:disulfide interchange protein dsbe (Cytochrome c biogenesis  
 DE protein cmcg).  
 GN DSBE OR CCMG OR H1095.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RX MEDLINE-95350630; PubMed-7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.E., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.",  
 RL Science 269:496-512(1995).  
 CC -1- FUNCTION: Involved in disulfide bond formation. Catalyzes a late,  
 CC reductive step in the assembly of periplasmic c-type cytochromes,  
 CC probably the reduction of disulfide bonds of the apocytochrome c  
 CC to allow covalent linkage with the heme. Possible subunit of a  
 CC heme lyase (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Mostly periplasmic; anchored in the inner  
 CC membrane (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. DSBE SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U32789; AAC22752.1; -  
 DR TIGR: H11095; -  
 DR InterPro: IPR004799; DSBE.  
 DR InterPro: IPR000063; ThioRed.  
 DR ProDom: PD003679; DSBE; 1.  
 DR TIGRFAMS: TIGR00385; dsbe; 1.  
 DR PROSITE: PS00194; THIOREDOXIN; 1.  
 KW Cytochrome c-type biogenesis; Redox-active center; Transmembrane;  
 KW Inner membrane; Complete proteome.  
 FT DOMAIN 1 4  
 FT TRANSMEM 5 23 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 24 181 POTENTIAL.  
 FT DISULFID 78 81 PERIPLASMIC (POTENTIAL).  
 FT DISULFID 81 81 REDOX-ACTIVE (BY SIMILARITY).  
 SQ SEQUENCE 181 AA: 20535 MW: 4B334B44179EF24 CRC64;  
 Query Match 1.3%; Score 7; DB 1; Length 181;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 426 TFIYGVK 432  
 |||||||  
 Db 144 TFIYGVK 150

```

RESULT 12
LPCA_HELPJ STANDARD: PRT: 192 AA.
AC 09ZKZ1:
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Phosphoglucose isomerase (EC 5.-.-.-).
GN LPCA OR GMAH OR JHP0791.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
ON NCBI_Taxid=85963;
RX SEQUENCE FROM N.A.
RA MEDLINE-99120557; PubMed-9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Mckelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -1- FUNCTION: INVOLVED IN SYNTHESIS OF GLYCEROMANNONEPTOSE 7-
CC PHOSPHATE (BY SIMILARITY).
CC -1- PATHWAY: Inner core lipopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. LPCA SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: AE001509; AAD06367.1; -
CC DR InterPro: IPR004515; GMAH.
CC DR InterPro: IPR001347; SIS.
CC DR Pfam: PF01380; SIS. 1.
CC DR TIGRGRAMS: TIGR00441; gmaH; 1.
CC KM Isomerase; Lipopolysaccharide biosynthesis; Complete proteome.
CC SQ SEQUENCE 192 AA; 21023 MW; 6270D17DC78C470A CRC64;

Query Match 1.3%; Score 7; DB 1; Length 192;
Best local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 326 LQDALKQ 332
DB 24 LQDALKQ 30

RESULT 13
YCSF_BACSU STANDARD: PRT: 211 AA.
ID YCSF_BACSU
AC P42963:
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Hypothetical protein ycsf.
GN YCSF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
ON NCBI_Taxid=1423;
RX SEQUENCE FROM N.A.
RA MEDLINE-168;
RA MEDLINE-97124189; PubMed-8969502;
RA Yamane K., Kumano M., Kurita K.;

```

```

RT RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A.M., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Chim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Melliado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Prescann E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche M., Sataie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiyuchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Tempstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Mambull R., Wedler E., Medler H., Weitzenger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumslein E., Yoshikawa H., Darchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 46-211 FROM N.A.
RX STRAIN-168;
RA Akagawa E., Kurita K., Sugawara T., Nakamura K., Yamane K.;
RA Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO E. NIDULANS LACTAM UTILIZATION PROTEIN LAMB AND
CC E. COLI YBGL AND H. INFLUENZAE HT1729.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: D50453; BAA09036.1; -
CC DR EMBL: D59106; CAB12213.1; -
CC DR EMBL: D38161; BAA07357.1; -
CC DR Subtilisin; BG11227; ycsf.
CC KM Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 211 AA; 22731 MW; 03660E1167B0C77E CRC64;

Query Match 1.3%; Score 7; DB 1; Length 211;
Best local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 122 LAERGV 128
DB 52 LAERGV 58

RESULT 14
PSPA_FCOLI

```

ID PSPA\_ECOLI STANDARD: PRT: 221 AA.  
 AC P23853; P76040; P77363;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phage shock protein A.  
 GN PSPA OR B1304 OR Z2482 OR ECS1881.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-K12;  
 RA Bristette J.L., Weiner L., Rimpaster T.L., Model P.;  
 RT "Characterization and sequence of the Escherichia coli stress-induced  
 RT psp operon.";  
 RL J Mol. Biol. 220:35-48(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RA MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12;  
 RA MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,  
 RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,  
 RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,  
 RA Motomura K., Nakase S., Nakamura Y., Nishimoto H., Nishio Y.,  
 RA Oshima H., Saito N., Sempel G., Seki Y., Sivasubraman S.,  
 RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,  
 RA Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RA MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / RIMD 0509952;  
 RA MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata K., Tanaka M., Tohe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kunara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [6]  
 RP PARTIAL SEQUENCE.  
 RA MEDLINE=95005446; PubMed=7921245;  
 RA Beugler H., Abraham D., Aschauer H., Turnowsky F.;  
 RT "Inhibition of lipid biosynthesis induces the expression of the pspa  
 RT gene.";

RL Microbiology 140:1937-1944(1994).  
 RN [7]  
 RP PARTIAL SEQUENCE OF 1-19.  
 RC STRAIN-K12 / JA300;  
 RA MEDLINE=96154424; PubMed=9493373;  
 RA Kobayashi H., Yamamoto M., Aono R.;  
 RT "Appearance of a stress-response protein, phage-shock protein A, in  
 RT Escherichia coli exposed to hydrophobic organic solvents.";  
 RL Microbiology 144:353-359(1998).  
 RN [8]  
 RP FUNCTION.  
 RA MEDLINE=96176771; PubMed=8598199;  
 RA Kleerebezem M., Crieland W., Tommassen J.;  
 RT "Involvement of stress protein Pspa (phage shock protein A) of  
 RT Escherichia coli in maintenance of the protonmotive force under  
 RT stress conditions.";  
 RL EMBO J. 15:162-171(1996).  
 CC -1- FUNCTION: THE PHAGE SHOCK PROTEIN (PSP) OPERON (PSPACE) MAY PLAY  
 CC A SIGNIFICANT ROLE IN THE COMPETITION FOR SURVIVAL UNDER NUTRIENT-  
 CC OR ENERGY-LIMITED CONDITIONS. PSPA HAS A ROLE AS A NEGATIVE  
 CC REGULATOR, IS REQUIRED FOR EFFICIENT TRANSLOCATION AND HAS AN  
 CC UNKNOWN FUNCTION IN THE MAINTENANCE OF THE PROTON MOTIVE FORCE.  
 CC -1- SUBCELLULAR LOCATION: Inner membrane-associated.  
 CC -1- INDUCTION: BY HEAT, ETHANOL, OSMOTIC SHOCK AND INFECTION BY  
 CC FILAMENTOUS BACTERIOPHAGES.  
 CC -1- SIMILARITY: BELONGS TO THE PSPA/TM30 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X57560; CAA40789.1; -  
 DR EMBL: AE000228; AAC74366.1; -  
 DR EMBL: D90768; BAA14873.1; -  
 DR EMBL: D90769; BAA14881.1; -  
 DR EMBL: AE005376; AAG56500.1; -  
 DR EMBL: AP002556; BAB35304.1; -  
 DR PIR: S17121; S17121.  
 DR EC02DBASE: D026.3; 6TH EDITION.  
 DR EC02DBASE: E026.0; 6TH EDITION.  
 DR Ecogene: E610776; pspa.  
 KW Coiled coil; Inner membrane; Complete proteome.  
 FT INIT\_MET 0  
 FT DOMAIN 29 186 COILED COIL (POTENTIAL).  
 FT COMBLIT 203 221 COILED COIL (POTENTIAL).  
 FT CONFLICT 181 181 A->R (IN REF. 1).  
 SQ SEQUENCE 221 AA: 25361 MW; 86F2C195702E5DCA CRC64;  
 Query Match 1.3%; Score 7; DB 1; Length 221;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 88 ARAALIE 94  
 DB 86 ARAALIE 92  
 RESULT 15  
 RL2\_SULSO  
 ID RL2\_SULSO STANDARD: PRT: 238 AA.  
 AC Q9UXA5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 50S ribosomal protein L2P.  
 GN RPL2P OR RPL2AB OR SS00716 OR C10\_014.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.

```

OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=20165948; PubMed=1071121;
RA Charlebois R.L., Singh R.K., Chan-Weher C.C.-Y., Allard G., Chow C.,
RA Cantaloniieri F., Curtis B., Duguet M., Erasuo G., Faguy D.,
RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA Kuschwah N., Lafleur E., Medina N., Peng X., Penny S.L., She O.,
RA St Jean A., Senen C.W., Young F., Zivanovic Y., Doilittle W.F.,
RA Ragan M.A., Erasuo G., Erasuo G., Erasuo G., Erasuo G., Erasuo G.,
RT "Gene content and organization of a 281-kbp contig from the genome of
RL the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";
RN Genome 43:116-136(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She O., Singh R.K., Cantaloniieri F., Zivanovic Y., Allard G.,
RA Aweez M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Charlebois R.L., Doilittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y18930; CAB57587.1; -
DR EMBL: AE006697; AAK41015.1; -
DR HSSP: P04257; 1RL2.
DR InterPro: IPR002171; Ribosomal_L2.
DR Pfam: PF00181; Ribosomal_L2.1.
DR PROSITE: PS00467; RIBOSOMAL_L2; FALSE-NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 238 AA; 25365 MW; B5CD0765610DC00 CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 238;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 KPLKAG 394
Db 166 KPLKAG 172

```

```

RT Aplysia synaptosomes.";
RN J. Neurosci. 21:382-391(2001).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN CONTROLLING CELL GROWTH
CC AND PROLIFERATION THROUGH THE SELECTIVE TRANSLATION OF PARTICULAR
CC CLASSES OF mRNA (BY SIMILARITY).
CC -1- PTM: Ribosomal protein S6 is the major substrate of protein
CC kinases in eukaryote ribosomes (By similarity).
CC -1- SIMILARITY: BELONGS TO THE S6E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF294917; AAG60623.1; -
DR InterPro: IPR001377; Ribosomal_S6E.
DR Pfam: PF01092; Ribosomal_S6E; 1.
DR ProDom: PD003460; Ribosomal_S6E; 1.
DR PROSITE: PS00578; RIBOSOMAL_S6E; 1.
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 247 AA; 28527 MW; 05A7D5B2FA5CBDF CRC64;

Query Match
Best Local Similarity 1.3%; Score 7; DB 1; Length 247;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 RVRLLD 135
Db 72 RVRLLD 78

RESULT 17
RL2_SCHPO STANDARD; PRT; 253 AA.
AC P08093; P14067; P36593; Q90U31;
DR 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L2 (K5) (K37) (K44).
GN (RPL2A OR RPL8-1 OR RPK5A OR RPK5 OR SPAC21E11.02 OR SPAC1F7.13C) AND
GN (RPL2B OR RPL8-2 OR RPK5B OR RPK37 OR SPBC2F12.07C OR SPBC24B9.04) AND
GN (RPL2C OR RPL8-3 OR RPK4 OR SPBC839.04).
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87146475; PubMed=3029717;
RA Nischt R., Gross T., Gattermann K.B., Swida U., Kaeuffer N.F.;
RT "Sequence and regulatory responses of a ribosomal protein gene from
RT the fission yeast Schizosaccharomycetes pombe."
RN Nucleic Acids Res. 15:1477-1492(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90124720; PubMed=2611912;
RA Gattermann K.B., Teletski C., Gross T., Kaeuffer N.F.;
RT "A ribosomal protein gene family from Schizosaccharomycetes pombe
RT consisting of three active members."
RN Curr. Genet. 16:361-367(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=90098792; PubMed=2602122;
RA Teletski C., Kaeuffer N.F.;
RT "Sequence of the ribosomal protein gene KD4 from Schizosaccharomycetes
RT pombe."
RN Nucleic Acids Res. 17:10118-10118(1989).
RN [4]

```

RP SEQUENCE FROM N.A. (RPL2A; RPL2B AND RPL2C).  
 RC STRAIN-972;  
 RX MEDLINE-21848401; PubMed-11859360;  
 RA Wood V., William R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Wellens J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs I., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,  
 RA Dada R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Domiguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Shpakovskii G.V., Usery D., Barrell B.G., Nurse P.;  
 RA "The genome sequence of *Schizosaccharomyces pombe*.";  
 RL Nature 415:871-880(2002).  
 RN [5]  
 RP SEQUENCE OF 1-47 FROM N.A.  
 RC STRAIN-968 h90;  
 RX MEDLINE-20223868; PubMed-10759889;  
 RA Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Harauchi T.,  
 RA Hiraoka Y.;  
 RT "Large-scale screening of intracellular protein localization in living  
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library.";  
 RL Genes Cells 5:169-190(2000).  
 CC -1- MISCELLANEOUS: THERE ARE THREE GENES FOR L2 IN S.POMBE.  
 CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts  
 CC in position 20 and 209.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, X05036; CAA28710.1; ALT\_FRAME.  
 DR EMBL, X51659; CAA35971.1;  
 DR EMBL, X16392; CAA34428.1;  
 DR EMBL, Z67999; CAA91962.1;  
 DR EMBL, Z67998; CAA91960.1;  
 DR EMBL, Z97211; CAB10155.1;  
 DR EMBL, AL096796; CAB46697.1;  
 DR EMBL, AB027845; BAA87149.1;  
 DR PIR, A48328; A48328.  
 DR PIR, S07377; S07377.  
 DR PIR, S06692; R5ZPD4.  
 DR PIR, B48328; B48328.  
 DR InterPro; IPR002171; Ribosomal\_L2.  
 DR Pfam; PF00181; Ribosomal\_L2; 1.  
 DR PROSITE; PS00467; RIBOSOMAL\_L2; 1.  
 KW Ribosomal protein; Multigene family.  
 FT CONFLICT 135 135 I -> S (IN REF. 2).  
 FT CONFLICT 158 158 P -> L (IN REF. 3).  
 SO SEQUENCE 253 AA; 27100 MW; 5FEB503EA376EA63 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 46;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 388 KPLKAG 394  
 |||||  
 Db 176 KPLKAG 182  
 RESULT 18  
 RL2 YEAST  
 ID RL2 YEAST STANDARD; PRT; 253 AA.  
 AC P05736;  
 DT 01-NOV-1998 (Rel. 09, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 60S ribosomal protein L2 (YL6) (U5) (RP8)  
 GN (RPL2A OR RPL5B OR YER031BC) AND (RPL2B OR RPL5A OR YIL018W).  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A. (RPL2B).  
 RC STRAIN-S288c / AB972;  
 RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Hornsby T.S., Hunt S., Jagels K., Jones M.,  
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.;  
 RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (RPL2A).  
 RC STRAIN-S288c / AB972;  
 RX MEDLINE-95400292; PubMed-7670463;  
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,  
 RA Sasamura S.-I., Sasamura M., Tsuchiya Y., Soeda E., Yokoyama K.,  
 RA Yamazaki M., Tashiro H., Eki T.;  
 RT "Analysis of the nucleotide sequence of chromosome VI from  
 RT *Saccharomyces cerevisiae*.";  
 RL Nat. Genet. 10:261-268(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (RPL2A).  
 RC STRAIN-S288c / AB972;  
 RX MEDLINE-96287654; PubMed-8686381;  
 RA Eki T., Naitou M., Hagiwara H., Abe M., Ozawa M., Sasamura S.-I.,  
 RA Sasamura M., Tsuchiya Y., Shibata T., Watanabe K., Ono A.,  
 RA Yamazaki M.-A., Tashiro H., Hanaoka F., Murakami Y.;  
 RT "Fifteen open reading frames in a 30.8 kb region of the right arm of  
 RT chromosome VI from *Saccharomyces cerevisiae*.";  
 RL Yeast 12:177-190(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (RPL2A AND RPL2B).  
 RC STRAIN-DBY746;  
 RX MEDLINE-95272535; PubMed-7753035;  
 RA Moore J., Jacobs H.T., Kaiser K.;  
 RT "Characterisation of *Saccharomyces cerevisiae* genes encoding  
 RT ribosomal protein YL6.";  
 RL Mol. Gen. Genet. 247:247-254(1995).  
 RN [5]  
 RP SEQUENCE OF 1-40.  
 RA Otake E., Higo K.-I., Itoh T.;  
 RT "Yeast ribosomal protein. VIII. Isolation of two proteins and  
 RT sequence characterization of twenty-four proteins from cytoplasmic  
 RT ribosomes.";  
 RL Mol. Gen. Genet. 195:544-546(1984).  
 RN [6]  
 RP SEQUENCE OF 1-20  
 RX MEDLINE-92184799; PubMed-1544921;  
 RA Takakura H., Tsunasawa S., Miyagi M., Warner J.R.;  
 RT "NH2-terminal acetylation of ribosomal proteins of *Saccharomyces  
 RT cerevisiae*.";  
 RL J. Biol. Chem. 267:5442-5445(1992).  
 RN [7]

RP SEQUENCE OF 225-253 FROM N.A. (RPL2A).  
 RX MEDLINE-95212908; PubMed-7698648;  
 RA Strunnikov A.V., Hogan E., Koshlund D.;  
 RT "SMC2, a Saccharomyces cerevisiae gene essential for chromosome  
 segregation and condensation, defines a subgroup within the SMC  
 family.";  
 RL Genes Dev. 9:587-599(1995).  
 CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L2 IN YEAST.  
 CC -1- SIMILARITY: BELONGS TO THE L2P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 CC  
 CD EMBL; 246881; CAA6974.1; -  
 DR EMBL; D50617; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL; U17359; AAA92283.1; -  
 DR EMBL; U17360; AAA92284.1; -  
 DR EMBL; U05820; AAA17418.1; -  
 DR PIR; S11254; S11254.  
 DR HSSP; P04257; 1RL2.  
 DR SGD; S0002104; RPL2A.  
 DR SGD; S0001280; RPL2B.  
 DR InterPro: IPR002171; Ribosomal\_L2.  
 DR Pfam: PF00181; Ribosomal\_L2.1.  
 DR PROSITE: PS00467; RIBOSOMAL\_L2.1.  
 KW Ribosomal protein; Multigene family.  
 FT INIT\_MEN 0  
 FT SEQUENCE 253 AA; 27277 MW; 51606080575FC95A CRC64;  
 SQ  
 Query Match 1.3%; Score 7; DB 1; Length 253;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 388 KPLKNG 394  
 |||||  
 Db 176 KPLKNG 182  
 RESULT 19  
 HA2R\_HUMAN  
 ID HA2R\_HUMAN STANDARD; PRT; 254 AA.  
 AC P01903;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE HLA class II histocompatibility antigen, DR alpha chain precursor.  
 GN HLA-DRA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 OX RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-83221632; PubMed-6304715;  
 RA Das H.K., Lawrence S.K., Weissman S.M.;  
 RT "Structure and nucleotide sequence of the heavy chain gene of  
 RT HLA-DR.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:3543-3547(1983).  
 RT [2]  
 RP REVISIONS.  
 RA Das H.K., Lawrence S.K., Weissman S.M.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7024-7024(1983).  
 RN [3]  
 RP SEQUENCE OF 26-204.  
 RX MEDLINE-82263347; PubMed-6955253;  
 RA Yang C.-Y., Kratzin H., Gatz H., Thines F.P., Kruse T., Egert G.,  
 RA Pauly E., Kolbel S., Weinet P., Hilschmann N.;  
 RT "Primary structure of class II human histocompatibility antigens. 2nd

RT Communication. Amino acid sequence of the N-terminal 179 residues of  
 RT the alpha-chain of an HLA-Dw2/DR2 alloantigen.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 363:671-676(1982).  
 RN [4]  
 RP SEQUENCE OF 26-60. AND SEQUENCE OF 32-202 AND 204-254 FROM N.A.  
 RX MEDLINE-83025073; PubMed-6812963;  
 RA Larhammar D., Gustafsson K., Claesson L., Bill P., Wiman K.,  
 RA Schenning L., Sundelin J., Widmark E., Peterson P.A., Rask L.;  
 RT "Alpha chain of HLA-DR transplantation antigens is a member of the  
 RT same protein superfamily as the immunoglobulins.";  
 RL Cell 30:153-161(1982).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-84057142; PubMed-6416803;  
 RA Kajimura Y., Toyoda H., Sato M., Miyakoshi S., Kaplan S.A., Ike Y.,  
 RA Goyert S.M., Silver J., Hawke D., Shively J.E., Suggs S.V.,  
 RA Wallace R.B., Itakura K.;  
 RT "Cloning the heavy chain of human HLA-DR antigen using synthetic  
 RT oligodeoxynucleotides as hybridization probes.";  
 RL DNA 2:175-182(1983).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-84169507; PubMed-6324094;  
 RA Schamboeck A., Korman A.J., Kamb A., Strominger J.L.;  
 RT "Organization of the transcriptional unit of a human class II  
 RT histocompatibility antigen: HLA-DR heavy chain.";  
 RL Nucleic Acids Res. 11:8663-8675(1983).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-83013020; PubMed-6811954;  
 RA Lee J.S., Trowsdale J., Travers P.J., Carey J., Grosveld F.,  
 RA Jenkins J., Bodmer W.F.;  
 RT "Sequence of an HLA-DR alpha-chain cDNA clone and Intron-exon  
 RT organization of the corresponding gene.";  
 RL Nature 299:750-752(1982).  
 RN [8]  
 RP SEQUENCE OF 29-254 FROM N.A.  
 RX MEDLINE-83299916; PubMed-6821129;  
 RA Korman A.J., Auffray C., Schamboeck A., Strominger J.L.;  
 RT "The amino acid sequence and gene organization of the heavy chain of  
 RT the HLA-DR antigen: homology to immunoglobulins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:6013-6017(1982).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91010755; PubMed-2212658;  
 RA Koppelman B., Cresswell P.;  
 RT "Rapid nonlysosomal degradation of assembled HLA class II  
 RT glycoproteins incorporating a mutant DR alpha-chain.";  
 RL J. Immunol. 145:2730-2736(1990).  
 RN [10]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 28-207.  
 RX MEDLINE-94195388; PubMed-8145819;  
 RA Stern L.J., Brown J.H., Jardtetzky T.J., Gorga J.C., Urban R.G.,  
 RA Strominger J.L., Wiley D.C.;  
 RT "Crystal structure of the human class II MHC protein HLA-DRI  
 RT complexed with an influenza virus peptide.";  
 RL Nature 368:215-221(1994).  
 RN [11]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 28-207.  
 RX MEDLINE-93302847; PubMed-8316295;  
 RA Brown J.H., Jardtetzky T.S., Gorga J.C., Stern L.J., Urban R.G.,  
 RA Strominger J.L., Wiley D.C.;  
 RT "Three-dimensional structure of the human class II histocompatibility  
 RT antigen HLA-DRI.";  
 RL Nature 364:33-39(1993).  
 RN [12]  
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF COMPLEX WITH SEB.  
 RX MEDLINE-94203282; PubMed-8152483;  
 RA Jardtetzky T.S., Brown J.H., Gorga J.C., Stern L.J., Urban R.G.,  
 RA Chi Y.I., Stauffer C., Strominger J.L., Wiley D.C.;  
 RT "Three-dimensional structure of a human class II histocompatibility  
 RT molecule complexed with superantigen.";  
 RL Nature 368:711-718(1994).

```

RN [13]
RP X-RAY CRYSTALLOGRAPHY (2.75 ANGSTROMS).
RX MEDLINE=96085023; PubMed=7477400;
RA Ghosh P., Amaya M., Mellins E., Wiley D.C.;
RT "The structure of an intermediate in class II MHC maturation: CLIP
RT bound to HLA-DR3."
RL Nature 378:457-462(1995).
RN [14]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF COMPLEX WITH COLLAGEN.
RX MEDLINE=98014591; PubMed=9354468;
RA Dessen A., Lawrence C.M., Cupo S., Zaller D.M., Wiley D.C.;
RT "X-ray crystal structure of HLA-DR4 (DRA*0101, DRB1*0401) complexed
RT with a peptide from human collagen II."
RL Immunity 7:473-481(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V00523; CAA23782.1; -
DR EMBL; J00204; AAA36302.1; -
DR EMBL; J00203; AAA36302.1; JOINED.
DR EMBL; X00274; CAA25076.1; ALT-INIT.
DR EMBL; J00194; AAA36275.1; -
DR EMBL; K01171; AAA59785.1; -
DR EMBL; M60334; AAA59783.1; -
DR PIR; A02206; HLHDA.
DR PDB; 1DLH; 22-JUN-94.
DR PDB; 1SEB; 20-JUN-96.
DR PDB; 2SEB; 28-JAN-98.
DR PDB; 1A0D; 28-JAN-98.
DR PDB; 1A6A; 27-MAY-98.
DR Genew; HGNC:4947; HLA-DRA.
DR MIM; 142860; -
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR001003; MHC_II_alpha.
DR Pfam; PF00047; Ig_1.
DR Pfam; PF00993; MHC_II_alpha; 1.
DR SMART; SM00407; IGc1; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 25
FT CHAIN 1 254
FT FT 26 254 HLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT FT 26 109 DR ALPHA CHAIN.
FT FT 110 203 EXTRACELLULAR ALPHA-1.
FT FT 204 213 EXTRACELLULAR ALPHA-2.
FT FT 217 239 CONNECTING PEPTIDE.
FT FT 240 254 CYTOPLASMIC.
FT FT 132 188 BY SIMILARITY.
FT FT 103 103 N-LINKED (GLCNAC. . .).
FT FT 143 143 N-LINKED (GLCNAC. . .).
FT FT 242 242 V->L.
FT FT 28 29 /FTID=VAR_004399.
FT FT 33 33 EE->AD (IN REF. 4).
FT FT 34 35 I->T (IN REF. 4).
FT FT 48 48 OA->YP (IN REF. 4).
FT FT 54 54 M->Q (IN REF. 4).
FT FT 67 67 D->T (IN REF. 4).
FT FT 67 67 V->A (IN REF. 3).
FT FT 149 149 V->E (IN REF. 3).
FT FT 254 254 N->E (IN REF. 3).
SQ SEQUENCE 254 AA; 28607 MW; 3CD1DBA952B2350 CRC64;
Query Match 1.3%; Score 7; DB 1; Length 254;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 364 VTVLINS 370

```

```

DB 114 VTVLINS 120
RESULT 20
Y683_HALNI
ID Y683_HALNI STANDARD; PRT; 263 AA.
AC 09HR12;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative aldolase Vng0683c (EC 4.2.1.-).
GN VNG0683C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Strogon J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welt R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danon M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angeline C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlenschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005014; AAG19176.1; -
DR InterPro; IPR002915; DEOC.
DR Pfam; PF01791; DEOC; 1.
KW Hypothetical protein; Lyase; Schiff base; Complete proteome.
FT BINDING 177 177 SCHIFF-BASE (BY SIMILARITY).
SQ SEQUENCE 263 AA; 28083 MW; 25FE044BA77494F CRC64;
Query Match 1.3%; Score 7; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 405 AVPATKD 411
DB 33 AVPATKD 39
RESULT 21
TSNX_HUMAN
ID TSNX_HUMAN STANDARD; PRT; 290 AA.
AC Q99598;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Translin-associated protein X (Translin-associated factor X).
GN TSNX OR TRAX.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Spleen;
MEDLINE=97165975; PubMed=9013868;

```



RA Aoki K., Ishida R., Kasai M.;  
 RT "Isolation and characterization of a cDNA encoding a translin-like  
 RT protein, TRAX.";  
 RL FEBS Lett. 401:109-112(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Guyonnet Duprat V., Dupuy D., Stef M., Arveiler B.;  
 RT "Genomic structure of human translin-associated factor X (TRAX)  
 RT gene.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain, and Muscle;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: INTERACTS WITH TRANSLIN.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: BELONGS TO THE TRANSLIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X95073; CAA64469.1; -;  
 CC DR EMBL: AF271269; AAK58640.1; -;  
 CC DR EMBL: AF271267; AAK58640.1; JOINED.  
 CC DR EMBL: AF271268; AAK58640.1; JOINED.  
 CC DR EMBL: BC010376; AAH10376.1; -;  
 CC DR EMBL: BC011797; AAH11797.1; -;  
 CC DR GeneW; HGNC:12380; TSNA.  
 CC DR MIM: 602964; -;  
 CC DR InterPro: IPR002848; Translin.  
 CC DR Pfam: PF01997; Translin; 1.  
 CC KW Nuclear protein.  
 CC SQ SEQUENCE 290 AA; 33112 MW; D3BBDD96B50C0B9 CRC64;  
 CC -----  
 CC Query Match 1.3%; Score 7; DB 1; Length 290;  
 CC Best Local Similarity 100.0%; Pred. No. 51;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QY 19 KTRSLIS 25  
 CC Db 138 KTRSLIS 144  
 CC -----  
 CC RESULT 22  
 CC BHCL\_RHOGO  
 CC ID BHCL\_RHOGO STANDARD; PRT; 291 AA.  
 CC AC P47231;  
 CC DT 01-FEB-1996 (Rel. 33, Created)  
 CC DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE Biphenyl-2,3-diol 1,2-dioxygenase I (EC 1.13.11.39) (23OHP  
 CC DE oxygenase I) (2,3-dihydroxybiphenyl dioxygenase I) (DHBD I).  
 CC GN BPHCI.  
 CC OS Rhodococcus globerulus.  
 CC OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 CC OC Actinomycetales; Corynebacteriaceae; Nocardiaceae; Rhodococcus.  
 CC OX NCBI\_TaxID=33008;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=P6;  
 CC RX MEDLINE=94171820; PubMed=8126007;  
 CC RA Asturias J.A., Ellis L.D., Frucha M., Timmis K.N.;  
 RT "Analysis of three 2,3-dihydroxybiphenyl 1,2-dioxygenases found in  
 RT Rhodococcus globerulus P6. Identification of a new family of  
 RT extradiol dioxygenases.";  
 RL J. Biol. Chem. 269:7807-7815(1994).  
 CC -1- CATALYTIC ACTIVITY: Biphenyl-2,3-di-ol + O(2) = 2-hydroxy-6-oxo-6-

CC phenylhexa-2,4-dienoate + H(2)O.  
 CC -1- COFACTOR: FERROUS ION.  
 CC -1- PATHWAY: DEGRADATION OF BIPIHENYLS AND POLYCHLOROBIPHENYLS (PCB) TO  
 CC BENZOIC ACID AND CHLOROBENZOIC ACIDS.  
 CC -1- SIMILARITY: BELONGS TO THE EXTRADIOL RING-CLEAVAGE DIOXYGENASE  
 CC FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X75633; CAA53297.1; -;  
 CC DR HSSP: P47228; 1HAN.  
 CC DR InterPro: IPR004360; Gly\_bleo\_diox.  
 CC DR InterPro: IPR000486; Xdiol\_dioxygenase.  
 CC DR Pfam: PF00903; Glyoxalase; 1.  
 CC DR ProDom: PD000977; Xdiol\_dioxygenase; 1.  
 CC DR PROSITE: PS00082; EXTRADIOL\_DIOXYGENAS; 1.  
 CC KW oxidoreductase; dioxygenase; Aromatic hydrocarbons catabolism; Iron.  
 CC FT METAL 146 146 IRON (BY SIMILARITY).  
 CC FT METAL 210 210 IRON (BY SIMILARITY).  
 CC FT METAL 260 260 IRON (BY SIMILARITY).  
 CC SQ SEQUENCE 291 AA; 32081 MW; 104F189FE1DDDA6A CRC64;  
 CC -----  
 CC Query Match 1.3%; Score 7; DB 1; Length 291;  
 CC Best Local Similarity 100.0%; Pred. No. 51;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QY 122 LAAERGV 128  
 CC Db 94 LAAERGV 100  
 CC -----  
 CC RESULT 23  
 CC Y766\_TREPA  
 CC ID Y766\_TREPA STANDARD; PRT; 293 AA.  
 CC AC O83747;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Hypothetical protein TP0766.  
 CC GN TP0766.  
 CC OS Treponema pallidum.  
 CC OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 CC OX NCBI\_TaxID=160;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=Nichols;  
 CC RX MEDLINE=98332770; PubMed=9665876;  
 CC RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,  
 RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,  
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,  
 RA Khalak H., Richardson P., Howell J.K., Chidambaram M., Utterback T.,  
 RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,  
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,  
 RA Venter J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 RT spirochete.";  
 RL Science 281:375-388(1998).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: AE001248; AAC65741.1; -;

DR TIGR: TP0766; -  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 293 AA: 32877 MW: F2091561B76AB196 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 293;  
Best Local Similarity 100.0%; Pred. No. 52;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 TSHALL 275  
DB 139 TSHALL 145

RESULT 24  
META\_VIBCH STANDARD; PRT; 313 AA.

AC O9KRM5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Homoserine O-succinyltransferase (EC 2.3.1.46) (Homoserine O-transuccinylase) (HTS).  
GN META OR VC1611.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Delgoutte I., Sellers P.,  
RA McDonald L., Utterback T., Fleischmann R.D., Mierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae".  
RL Nature 406:477-483(2000).  
CC -1- CATALYTIC ACTIVITY: Succinyl-CoA + L-homoserine -> CoA + O-succinyl-L-homoserine.  
CC -1- PATHWAY: Methionine biosynthesis; HTS variant; first step.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE HTS FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AE004238; AAF94765.1; ALT\_INT.  
DR TIGR: VC1611; -  
DR TIGR: TIGR01001; meta: 1.  
KW Methionine biosynthesis; Transferase; Acyltransferase;  
KW Complete proteome.  
FT ACT\_SITE 142  
SQ SEQUENCE 313 AA: 36199 MW: 9F77F1FBFF06DA89 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 313;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 DDIAT 224  
DB 207 DDIAT 213

RESULT 25  
Y486\_METJA

ID Y486\_METJA STANDARD; PRT; 322 AA.  
AC O57910;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein M00486.  
GN M00486.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii".  
RL Science 273:1058-1073(1996).  
CC -1- SIMILARITY: TO E.COLI YHCC.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U67498; AAB98477.1; -  
DR TIGR: M00486; -  
DR TIGR: TIGR01212; TIGR01212; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 322 AA: 37390 MW: F3D074B32180B805 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 322;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 513 AKILSL 519  
DB 218 AKILSL 224

RESULT 26  
E2B\_XYLFA STANDARD; PRT; 354 AA.  
AC O9PAG5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Putative translation initiation factor eIF-2B (eIF-2B GDP-GTP exchange factor).  
GN XE2553.  
OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OC Xylella.  
OX NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=945C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,  
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,

RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Faciniani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garter M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kurume E.E., Laigret F., Lamais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Neto L.E.S.,  
 RA Nhami A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peloto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E.Jr., de Sa R.G., Saneilli R.V., Sasaaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,  
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,  
 RA de Souza A.P., Terezi M.F., Trufi D., Tsai S.M., Tsubako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovskij-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.,  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"  
 CC Nature 406:151-159(2000).  
 CC -1- FUNCTION: CATALYZES THE EXCHANGE OF INITIATION FACTOR 2-BOUND GDP  
 CC FOR GTP (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EIF-2B ALPHA/BETA/DELTA SUBUNITS  
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: AE004062; AAF85350.1; -  
 DR InterPro: IPR000649; IF-2B.  
 DR Pfam: PF01008; IF-2B; 1.  
 DR TIGRfam: TIGR00524; eif-2B-rel; 1.  
 KM Hypothetical protein; Initiation factor; Protein biosynthesis;  
 KW Complete proteome.  
 SQ SEQUENCE 354 AA; 37845 MW; 5EAF417C919FCB76 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 359 QDCIDVT 365  
 DB 215 QDCIDVT 221

RESULT 27  
 SRDI CAEEL  
 AC 019992;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Serpentine receptor class delta 1 (Srd-1 protein).  
 GN SRD-1 OR F33H1.5.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Wilkinson J.;  
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE C.ELEGANS RECEPTOR-LIKE PROTEIN SRD  
 CC FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: Z48783; CAA8700.1; -  
 DR Wormpep: F33H1.5; CE02207.  
 DR InterPro: IPR003002; 7TM\_nematode.  
 DR InterPro: IPR000168; 7TM\_nematode.  
 DR InterPro: IPR000276; GPCR\_Rhodopsin.  
 DR Pfam: PF01461; 7tm\_4; 1.  
 DR PROSITE: PS50262; G-PROTEIN-RECEP\_FL2; UNKNOWN\_1.  
 KW Transmembrane; Multigene family.  
 FT TRANSMEM 31  
 FT TRANSMEM 62  
 FT TRANSMEM 82  
 FT TRANSMEM 109  
 FT TRANSMEM 129  
 FT TRANSMEM 148  
 FT TRANSMEM 168  
 FT TRANSMEM 209  
 FT TRANSMEM 229  
 FT TRANSMEM 267  
 FT TRANSMEM 287  
 FT TRANSMEM 295  
 FT TRANSMEM 315  
 SQ SEQUENCE 371 AA; 41927 MW; 186B1019833D2384 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 89 RAALIES 95  
 DB 11 RAALIES 17

RESULT 28  
 IDHA CAEEL  
 AC 093714;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial  
 DE precursor (EC 1.1.1.41) (isocitric dehydrogenase) (NAD-specific  
 DE ICDH).  
 GN F43G9.1.  
 OS Caenorhabditis elegans.  
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN 11  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Kershaw J.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN 12  
 RP REVISIONS.  
 RA Jones S.J.M.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: Isocitrate + NAD(+) = 2-oxoglutarate + CO(2) +  
 CC NADH.  
 CC -1- SUBUNIT: HETEROOLIGOMER OF SUBUNITS ALPHA, BETA, AND GAMMA IN THE  
 CC APPARENT RATIO OF 2:1:1 (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE  
 CC DEHYDROGENASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: 279755; CAB02111.1; .  
 DR HSSP: P08200; IISO.  
 DR WormRep: F4369.1; CE10354.  
 DR InterPro: IPR001804; Isoch.  
 DR InterPro: IPR004434; Mito\_nad\_idh.  
 DR Pfam: PF00180; Isoch. 1.  
 DR TIGRFAMS: TIGR00175; mito\_nad\_idh.  
 DR PROSITE: PS00470; IDH.IMDH. 1.  
 KW Oxidoreductase; NAD; Tricarboxylic acid cycle; Transit peptide;  
 KM Mitochondrion.  
 FT TRANSIT 1 ?  
 FT CHAIN ? 374  
 FT ACT\_SITE 102 102  
 FT BINDING 71EFDPE1FE2C9FD5 CRC64;  
 FT SEQUENCE 374 AA; 40509 MW; 71EFDPE1FE2C9FD5 CRC64;  
 Query Match  
 Best Local Similarity 100.0%; Score 7; DB 1; Length 374;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 SGNIGK 259  
 DB 268 SGNIGK 274

RESULT 29  
 GBB3\_TOBAC STANDARD; PRT; 375 AA.  
 ID Q40507;  
 AC 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Guanine nucleotide-binding protein beta subunit.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; eusterids I; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Samsun NN; TISSUE=leaf;  
 RA Kuznetsov V.V., Oelmüller R.;  
 RT "Isolation and characterization of cDNAs encoding the subunit beta of  
 heterotrimeric G proteins from N. tabacum.";  
 RL (in) Plant Gene Register PG96-048.  
 CC -I- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
 INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 EFFECTOR INTERACTION.  
 CC -I- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -I- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X98161; CAA66842.1; .  
 DR HSSP: P04901; TRBG.  
 DR InterPro: IPR001632; Gproteins\_B.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF004400; WD40. 7.  
 DR PRINTS: PR00319; GPROTEINB.  
 DR PRINTS: PR00320; GPROTEINBPT.  
 DR ProDom: PD000018; WD40; 5.

DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE: PS00682; WD\_REPEATS\_2; 5.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KW Transducer; Repeat; WD repeat; Multigene family.  
 FT REPEAT 63 93  
 FT REPEAT 105 135  
 FT REPEAT 154 185  
 FT REPEAT 202 233  
 FT REPEAT 246 276  
 FT REPEAT 293 323  
 FT REPEAT 339 369  
 FT SEQUENCE 375 AA; 40812 MW; 72D53D1DFC5CF598 CRC64;  
 Query Match  
 Best Local Similarity 100.0%; Score 7; DB 1; Length 375;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLLF 117  
 DB 305 ISGRLLF 311

RESULT 30  
 VATC\_ARATH STANDARD; PRT; 375 AA.  
 ID VATC\_ARATH  
 AC Q9SD57;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Vacuolar ATP synthase subunit C (EC 3.6.3.14) (V-ATPase C subunit)  
 DE (vacuolar proton pump C subunit).  
 GN VATC OR DET3 OR AT1G12840 OR F13K23.9.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA MEDLINE=20085001; PubMed=10617574;  
 RA Schumacher K., Vaeleod D., McCarthy M., Sze H., Wilkins T., Chory J.;  
 RT "The Arabidopsis delts mutant reveals a central role for the vacuolar  
 H(+) ATPase in plant growth and development.";  
 RL Genes Dev. 13:3259-3270(1999).  
 CC -I- FUNCTION: SUBUNIT OF THE PERIPHERAL V1 COMPLEX OF VACUOLAR ATPASE.  
 SUBUNIT C IS NECESSARY FOR THE ASSEMBLY OF THE CATALYTIC SECTOR OF  
 THE ENZYME AND IS LIKELY TO HAVE A SPECIFIC FUNCTION IN ITS  
 CATALYTIC ACTIVITY. V-ATPASE IS RESPONSIBLE FOR ACIDIFYING A

CC VARIETY OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(in) = ADP + phosphate +  
 CC H(+)(out).  
 CC -1- SUBUNIT: V-ATPase is a heteromultimeric enzyme composed of a  
 CC peripheral catalytic V1 complex (components A to H) attached to a  
 CC an integral membrane V0 proton pore complex (components: a, c, c',  
 CC c'', and d) (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE V-ATPASE C SUBUNIT FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF208261; AAF20146.1; -  
 DR EMBL: AC012187; AAF78489.1; -  
 DR InterPro: IPR004907; V-ATPase\_C.  
 DR Pfam: PF03223; V-ATPase\_C; 1.  
 KM Hydrolase; ATP synthesis; Hydrogen ion transport.  
 SO SEQUENCE 375 AA; 42619 MW; A50F11D168F31C97 CRC64;

Query Match 1.38; Score 7; DB 1; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 354 LAKYOD 360  
 |||||  
 DB 264 LAKYOD 270

RESULT 31  
 GB1 TOBAC STANDARD; PRT; 377 AA.  
 ID GB1 TOBAC  
 AC P93357;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Guanine nucleotide-binding protein beta subunit 1.  
 OS Nicotiana tabacum (Common tobacco)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanales;  
 OC NCBI\_TaxID=4097;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. SRL; TISSUE=Leaf;  
 RA Lein W., Saalbach G.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEIN (G PROTEINS) ARE  
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 CC EFFECTOR INTERACTION.  
 CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z84820; CAB06618.1; -  
 DR HSSP: P04901; ITBG.  
 DR InterPro: IPR001632; Gprotein\_B.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 7.  
 DR PRINTS: PR00319; GPROTEINB.  
 DR PRINTS: PR00320; GPROTEINBPT.  
 DR PRODOM: PD000018; WD40; 5.  
 DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 5.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KM Transducer; Repeat; WD repeat; Multigene family.  
 FT REPEAT 63 93  
 FT REPEAT 105 135  
 FT REPEAT 154 185  
 FT REPEAT 202 233  
 FT REPEAT 246 276  
 FT REPEAT 293 323  
 FT REPEAT 339 369  
 SO SEQUENCE 377 AA; 41145 MW; 23519B6467A079B CRC64;

DR PRINTS: PR00320; GPROTEINBPT.  
 DR PRODOM: PD000018; WD40; 5.  
 DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 5.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KM Transducer; Repeat; WD repeat; Multigene family.  
 FT REPEAT 63 93  
 FT REPEAT 105 135  
 FT REPEAT 154 185  
 FT REPEAT 202 233  
 FT REPEAT 246 276  
 FT REPEAT 293 323  
 FT REPEAT 339 369  
 SO SEQUENCE 377 AA; 41145 MW; 23519B6467A079B CRC64;

Query Match 1.38; Score 7; DB 1; Length 377;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLLF 117  
 |||||  
 DB 305 ISGRLLF 311

RESULT 32  
 GB2 TOBAC STANDARD; PRT; 377 AA.  
 ID GB2 TOBAC  
 AC P93358;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Guanine nucleotide-binding protein beta subunit 2.  
 OS Nicotiana tabacum (Common tobacco)  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanales; Solanales; Solanales;  
 OC NCBI\_TaxID=4097;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. SRL; TISSUE=Leaf;  
 RA Lein W., Saalbach G.;  
 RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEIN (G PROTEINS) ARE  
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 CC EFFECTOR INTERACTION.  
 CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z84821; CAB06619.1; -  
 DR HSSP: P04901; ITBG.  
 DR InterPro: IPR001632; Gprotein\_B.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 7.  
 DR PRINTS: PR00319; GPROTEINB.  
 DR PRINTS: PR00320; GPROTEINBPT.  
 DR PRODOM: PD000018; WD40; 5.  
 DR SMART: SM00320; WD40; 7.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 2.  
 DR PROSITE: PS50082; WD\_REPEATS\_2; 5.  
 DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
 KM Transducer; Repeat; WD repeat; Multigene family.  
 FT REPEAT 63 93  
 FT REPEAT 105 135  
 FT REPEAT 154 185  
 FT REPEAT 202 233  
 FT REPEAT 246 276  
 FT REPEAT 293 323  
 FT REPEAT 339 369  
 SO SEQUENCE 377 AA; 41145 MW; 23519B6467A079B CRC64;

```

FT REPEAT 105 135 WD 2.
FT REPEAT 154 185 WD 3.
FT REPEAT 202 233 WD 4.
FT REPEAT 246 276 WD 5.
FT REPEAT 293 323 WD 6.
FT REPEAT 339 369 WD 7.
SQ SEQUENCE 377 AA: 41084 MW: 28P2EB42472D3472 CRC64:
Query Match 1.3%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY 111 ISGRLEF 117
      |||||
Db 305 ISGRLEF 311

RESULT 33
GBB_NICPL STANDARD: PRT; 377 AA.
ID GBB_NICPL
AC P93339;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit (NGPBI).
OS Nicotiana glauca
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
[]
RP SEQUENCE FROM N.A.
RX MEDLINE=20225456; PubMed=10760577;
RA Kaydanov C., Tewes A., Adler K., Mantoufeli R.;
RT "Molecular characterization of cDNAs encoding G protein alpha and
RT beta subunits and study of their temporal and spatial expression
RT patterns in Nicotiana glauca."
RL Blochim. Biophys. Acta 1491:143-160(2000).
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC EFFECTOR INTERACTION.
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -!- SUBCELLULAR LOCATION: In plasma and endoplasmic reticulum
CC membranes.
CC -!- INDUCTION: By auxin.
CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on lists
CC modified by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-ch).
CC -----
DR EMBL: Y09513; CAAT0704.1; -
DR HSP, P04901; 1CG2.
DR InterPro: IPR001632; Gprotein_B.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 7.
DR PRINTS: PR00319; GPROTEINB.
DR PRINTS: PR00320; GPROTEINBPT.
DR PRODOM: PD000018; WD40; 5.
DR SMART: SM00320; WD40; 7.
DR PROSITE: PS00678; WD_REPEATS_1; 2.
DR PROSITE: PS00682; WD_REPEATS_2; 5.
DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
KW Transducer; Repeat; WD Repeat.
FT REPEAT 63 93 WD 1.
FT REPEAT 105 135 WD 2.
FT REPEAT 154 185 WD 3.

```

```

FT REPEAT 202 233 WD 4.
FT REPEAT 246 276 WD 5.
FT REPEAT 293 323 WD 6.
FT REPEAT 339 369 WD 7.
SQ SEQUENCE 377 AA; 41317 MW; 0351A96C73B511C5 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLLF 117
      |||||
Db 305 ISGRLLF 311

RESULT 34
GBB_SOLTU
ID GBB_SOLTU STANDARD. PRT. 377 AA.
AC P93563;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Guanine nucleotide-binding protein beta subunit.
GN GBI.
OC Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
XX
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Desiree;
RA Provant N.J.; Ma H.; Willmitzer L.; Mueller-Roeber B.;
RL submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
CC EFFECTOR INTERACTION.
CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb-sib.ch).
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL; X87837; CAA61106.1; -.
DR HSSP; P04901; I6G2.
DR InterPro; IPR001632; Gprotein_B.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 7.
DR PRINTS; PR00319; GPROTEINB.
DR PRINTS; PR00320; GPROTEINBPT.
DR PRODom; PD000018; WD40; 5.
DR SMART; SM00320; WD40; 7.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PSS0082; WD_REPEATS_2; 5.
DR PROSITE; PSS0294; WD_REPEATS_REGION; 1.
KW Transducer; Repeat; WD repeat.
FT REPEAT 63 93 WD 1.
FT REPEAT 105 135 WD 2.
FT REPEAT 154 185 WD 3.
FT REPEAT 202 233 WD 4.
FT REPEAT 246 276 WD 5.
FT REPEAT 293 323 WD 6.
FT REPEAT 339 369 WD 7.
SQ SEQUENCE 377 AA; 40941 MW; B9AFLAAC721DCBC1 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 377;

```

Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLEF 117  
|||||||  
Db 305 ISGRLEF 311

RESULT 35

GBB\_MAIZE STANDARD; PRT; 380 AA.

AC P49178;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Guanine nucleotide-binding protein beta subunit.

GN GBL.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;

OC Panicoideae; Andropogoneae; Zea.

OX NCBI\_TaxID=4577;

RN [1]

RP MEDLINE-95023950; PubMed-7937804;

RA Weiss C.A., Garnat C.W., Mukai K., Hu Y., Ma H.;

RT "Isolation of cDNAs encoding guanine nucleotide-binding protein beta-

RT subunit homologues from maize (ZGB1) and Arabidopsis (AGB1).";

CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE

CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE

CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE

CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-

CC EFFECTOR INTERACTION.

CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).

CC -1- TISSUE SPECIFICITY: PRESENT IN THE ROOT, LEAF, AND TASSEL.

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U12233; AAA50446.1; -

DR HSSP: P04901; ITBG.

DR MaizeDB: 105669; -

DR InterPro: IPR001632; Gprotein\_B.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00400; WD40; 7.

DR PRINTS: PR00319; GPROTEINB.

DR PRINTS: PR00320; GPROTEINBPT.

DR ProDom: PD000018; WD40; 5.

DR SMART: SM00320; WD40; 7.

DR PROSITE: PS00678; WD\_REPEATS\_1; 3.

DR PROSITE: PS00682; WD\_REPEATS\_2; 5.

DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.

KW Transducer; Repeat; WD repeat.

FT REPEAT 64 94 WD 1.

FT REPEAT 106 136 WD 2.

FT REPEAT 155 186 WD 3.

FT REPEAT 203 234 WD 4.

FT REPEAT 247 277 WD 5.

FT REPEAT 296 326 WD 6.

FT REPEAT 342 372 WD 7.

SEQUENCE 380 AA; 41714 MW; 7D5DC5276C33DD1E CRC64;

Query Match 1.3%; Score 7; DB 1; Length 380;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLEF 117  
|||||||  
Db 308 ISGRLEF 314

RESULT 36

GBB\_ORYZA

AC 040687; STANDARD; PRT; 380 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE Guanine nucleotide-binding protein beta subunit.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzaceae; Oryza.

OX NCBI\_TaxID=4530;

RN [1]

RP MEDLINE-96213007; PubMed-8665097;

RA Ishikawa A., Iwasaki Y., Asahi T.;

RT "Molecular cloning and characterization of a cDNA for the beta

RT subunit of a G protein from rice.";

RT Plant Cell Physiol. 37:223-228(1996).

CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE

CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE

CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE

CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-

CC EFFECTOR INTERACTION.

CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).

CC -1- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: X89737; CA61889.1; -

DR HSSP: P04901; ITBG.

DR InterPro: IPR001632; Gprotein\_B.

DR InterPro: IPR001680; WD40.

DR Pfam: PF00400; WD40; 7.

DR PRINTS: PR00319; GPROTEINB.

DR PRINTS: PR00320; GPROTEINBPT.

DR ProDom: PD000018; WD40; 4.

DR SMART: SM00320; WD40; 7.

DR PROSITE: PS00678; WD\_REPEATS\_1; 3.

DR PROSITE: PS00682; WD\_REPEATS\_2; 5.

DR PROSITE: PS50294; WD\_REPEATS\_REGION; 1.

KW Transducer; Repeat; WD repeat.

FT REPEAT 64 94 WD 1.

FT REPEAT 106 136 WD 2.

FT REPEAT 155 186 WD 3.

FT REPEAT 203 234 WD 4.

FT REPEAT 247 277 WD 5.

FT REPEAT 296 326 WD 6.

FT REPEAT 342 372 WD 7.

SEQUENCE 380 AA; 41751 MW; 36CE374E2A3E8EB0 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 380;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 ISGRLEF 117  
|||||||  
Db 308 ISGRLEF 314

```

RESULT 37
SCHE_STRHA          STANDARD:          PRT:          380 AA.
ID      SCHE_STRHA          STANDARD:          PRT:          380 AA.
AC      Q05361;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      01-FEB-1995 (Rel. 31, Last annotation update)
DE      SCHE protein.
GN      SCHE.
OS      Streptomyces halstedii.
OC      Bacteria; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Streptomycetaceae; Streptomycetes.
OX      NCBI_TaxID=1944;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93345807; PubMed=8344517;
RA      Blanco G., Brian P., Pereda A., Mendez C., Salas J.A., Chater K.F.;
RT      "Hybridization and DNA sequence analyses suggest an early
RT      evolutionary divergence of related biosynthetic gene sets encoding
RT      polyketide antibiotics and spore pigments in Streptomyces spp.";
RL      Gene 130:107-116(1993).
CC      -1- SIMILARITY: BELONGS TO THE SCHE/CURD FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L05390; AAA02831.1;
DR      PIR; J00823; J00823
SQ      SEQUENCE 380 AA; 41212 MW; 6F7C6F09D3EE9E6 CRC64;

Query Match          1.3%; Score 7; DB 1; Length 380;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      442 DPRSARL 448
DB      234 DPRSARL 240

RESULT 38
ISG2_STRCO          STANDARD:          PRT:          385 AA.
ID      ISG2_STRCO          STANDARD:          PRT:          385 AA.
AC      Q9YXR9;
DT      15-JUN-2002 (Rel. 41, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase 2.
GN      ISG2 OR GCPE2 OR SC05696 OR SC5H4.20.
OS      Streptomyces coelicolor.
OC      Bacteria; Actinobacteridae; Actinobacteridae;
OC      Actinomycetales; Streptomycetaceae; Streptomycetes.
OX      NCBI_TaxID=1902;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN=A3(2) / M145;
RX      MEDLINE=21996410; PubMed=12000953;
RA      Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA      Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA      Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA      Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA      Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA      Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA      Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA      Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA      Hopwood D.A.;
RT      "Complete genome sequence of the model actinomycete Streptomyces
RT      coelicolor A3(2).";
RL      Nature 417:141-147(2002).

```

```

CC      -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (By similarity).
CC      -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC      -1- SIMILARITY: BELONGS TO THE ISRG FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AL355913; CAB91132.1;
DR      InterPro; IPR004588; GCPE.
DR      TrifRAMS; TIGR00612; gcpe; 1.
KW      Isoprene biosynthesis; Complete proteome.
SQ      SEQUENCE 385 AA; 40795 MW; C69D4346FEAD5F7E CRC64;

Query Match          1.3%; Score 7; DB 1; Length 385;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      460 KIAE0ME 466
DB      364 KIAE0ME 370

RESULT 39
NR01_HUMAN          STANDARD:          PRT:          394 AA.
ID      NR01_HUMAN          STANDARD:          PRT:          394 AA.
AC      Q92597; O15207; Q9NTR6;
DT      15-JUL-1999 (Rel. 38, Created)
DT      15-JUL-1999 (Rel. 38, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      NDRG1 protein (N-myc downstream regulated gene 1 protein)
DE      (differentiation-related gene 1 protein) (DRG1) (Reducing agents and
DE      tunicamycin-responsive protein) (RTP) (Nickel-specific induction
DE      protein Cap43) (Rit42).
GN      NDRG1 OR RTP OR DRG1 OR CAP43.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      TISSUE=Umbilical vein endothelial cells;
RX      MEDLINE=97094664; PubMed=8939898;
RA      Kokame K., Kato H., Miyata T.;
RT      "Homocysteine-responsive genes in vascular endothelial cells
RT      identified by differential display analysis. GPR78/BIP and novel
RT      genes.";
RL      J. Biol. Chem. 271:29659-29665(1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97395543; PubMed=9251681;
RA      van Belzen N., Dinjens W.N.M., Diesveld M.P.G., Groen N.A.,
RA      van der Made A.C.J., Nozawa Y., Vlietstra R., Trapman J.,
RA      Bosman F.T.;
RT      "A novel gene which is up-regulated during colon epithelial cell
RT      differentiation and down-regulated in colorectal neoplasms.";
RL      Lab. Invest. 77:85-92(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      TISSUE=Lung;
RX      MEDLINE=98266801; PubMed=9605764;
RA      Zhou D., Salnikow K., Costa M.;
RT      "Cap43, a novel gene specifically induced by Ni2+ compounds.";
RL      Cancer Res. 58:2182-2189(1998).
RN      [4]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=99326165; PubMed=10395947;

```



RA Piquemal D., Jouila D., Balaguer P., Basset A., Marti J., Commes T.;  
 RT "Differential expression of the RTP/Digl/Ndr1 gene product in  
 RT proliferating and growth arrested cells";  
 RL Biochim. Biophys. Acta 1450:364-373(1999).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Kidney;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.  
 RN [6]  
 RP SEQUENCE OF 1-198 FROM N.A.  
 RA Bleischmidt K., Kalaydjieva L., Goodman R., Gresham D., Baas F.,  
 RA Jonge R.D., Schilhabel M., Schatevov R., Baumgart C., Menzel U.,  
 RA Rosenthal A.;  
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC TISSUE-Brain;  
 RA Angelicheva D., Kalaydjieva L.;  
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.  
 RN [8]  
 RP FUNCTION.  
 RX MEDLINE-98438066; PubMed-9766676;  
 RA Kurdistan S.K., Arizli P., Reimer C.L., Sugrue M.M., Aaronson S.A.,  
 RA Lee S.W.;  
 RT "Inhibition of tumor cell growth by RTP/rit42 and its responsiveness  
 to p53 and DNA damage.";  
 RL Cancer Res. 58:4439-4444(1998).  
 CC -1- FUNCTION: MAY HAVE A GROWTH INHIBITORY ROLE.  
 CC -1- SUBCELLULAR LOCATION: WHEREAS IN PROSTATE EPITHELIUM AND PLACENTAL  
 CC CHORION IT IS LOCATED IN BOTH THE CYTOPLASM AND THE NUCLEUS,  
 CC NUCLEAR STAINING IS NOT OBSERVED IN COLON EPITHELIUM CELLS.  
 CC INSTEAD ITS LOCALIZATION CHANGES FROM THE CYTOPLASM TO THE PLASMA  
 CC MEMBRANE DURING DIFFERENTIATION OF COLON CARCINOMA CELL LINES IN  
 CC VITRO.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOST PROMINENTLY IN  
 CC PLACENTAL MEMBRANES AND PROSTATE, KIDNEY, SMALL INTESTINE, AND  
 CC OVARY TISSUES. REDUCED EXPRESSION IN ADENOCARCINOMAS COMPARED TO  
 CC NORMAL TISSUES. IN COLON, PROSTATE AND PLACENTAL MEMBRANES, THE  
 CC CELLS THAT BORDER THE LUMEN SHOW THE HIGHEST EXPRESSION.  
 CC -1- INDUCTION: BY HOMOCYSTEINE, 2-MERCAPTOETHANOL, TUNICAMYCIN IN  
 CC ENDOTHELIAL CELLS. INDUCED APPROXIMATELY 20-FOLD DURING IN VITRO  
 CC DIFFERENTIATION OF THE COLON CARCINOMA CELL LINES HT29-D4 AND  
 CC CACO-2. INDUCED BY NICKEL COMPOUNDS IN ALL TESTED CELL LINES. THE  
 CC PRIMARY SIGNAL FOR ITS INDUCTION IS AN ELEVATION OF FREE  
 CC INTRACELLULAR CALCIUM ION CAUSED BY NICKEL ION EXPOSURE. OKADAIC  
 CC ACID, A SERINE/THREONINE PHOSPHATASE INHIBITOR, INDUCED ITS  
 CC EXPRESSION MORE RAPIDLY AND MORE EFFICIENTLY THAN NICKEL.  
 CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D87953; BAA13505.1; -  
 DR EMBL: X82845; CAA63430.1; -  
 DR EMBL: AF004162; AAC13419.1; -  
 DR EMBL: AF186190; AAG29875.1; -  
 DR EMBL: AF186190; AAF00622.1; -  
 DR EMBL: BC003175; AAH03175.1; -  
 DR EMBL: AF230380; AAF71305.1; -  
 DR Genew: HGNC:7679; NDRG1.  
 DR MIM: 605262; -  
 DR InterPro: IPR004142; Ndr.  
 DR InterPro: IPR000379; Ser\_estrs\_site.  
 DR Pfam: PF03096; Ndr; 1.  
 KM Nuclear protein: Repeat.  
 FT DOMAIN 339 368 3 X 10 AA TANDEM REPEATS OF G-T-R-S-R-S-  
 FT H-T-S-E.

FT REPEAT 339 348 1.  
 FT REPEAT 349 358 2.  
 FT REPEAT 359 368 3.  
 FT CONFLICT 145 145 I -> T (IN REF. 2).  
 SQ SEQUENCE 394 AA; 42835 MW; 4C816B9C85E3756F CRC64;  
 Query Match 1.3%; Score 7; DB 1;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 416 GSSVTSL 422  
 Db 331 GSSVTSL 337  
 RESULT 40  
 ID NDR1\_MOUSE STANDARD; PRT; 394 AA.  
 AC Q62433; P97862;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NDRG1 protein (N-myc downstream regulated gene 1 protein) (Protein  
 DE Ndr1).  
 GN NDRG1 OR NDR1 OR NDRL OR TDD5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-99310663; PubMed-10381566;  
 RA Shimono A., Okuda T., Kondoh H.;  
 RT "N-myc-dependent repression of ndr1, a gene identified by direct  
 RT subtraction of whole mouse embryo cDNAs between wild type and N-myc  
 RT mutant.";  
 RL Mech. Dev. 83:39-52(1999).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Hybridoma;  
 RX MEDLINE-97289706; PubMed-9144177;  
 RA Lin T.-M., Chang C.;  
 RT "Cloning and characterization of TDD5, an androgen target gene that is  
 RT differentially repressed by testosterone and dihydrotestosterone.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:4988-4993(1997).  
 CC -1- FUNCTION: MAY HAVE A GROWTH INHIBITORY ROLE (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN KIDNEY FOLLOWED  
 CC BY NERVOUS TISSUES (HYPOPHALAMUS, CEREBELLUM, AND CEREBRUM) AND  
 CC PREPUTIAL GLAND, AND IN MUCH SMALLER QUANTITIES IN OTHER TISSUES.  
 CC EXCEPT DUODENUM AND PROSTATE.  
 CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION INCREASES AFTER 13.5 DPC WHEN  
 CC N-MYC EXPRESSION IS LARGELY DOWN-REGULATED.  
 CC -1- INDUCTION: REPRESSED BY TESTOSTERONE AND ALSO IN A LESSER EXTENT  
 CC BY DIHYDROTESTOSTERONE.  
 CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U60593; AAB03484.1; -  
 DR EMBL: U52073; AAB58249.1; -  
 DR MGD: MGI:1341799; Ndr1.  
 DR InterPro: IPR004142; Ndr.  
 DR Pfam: PF03096; Ndr; 1.  
 KM Nuclear protein: Repeat.  
 FT DOMAIN 339 368 3 X 10 AA TANDEM REPEATS OF G-[PST]-R-S-  
 FT R-S-H-T-S-E.

```
FT REPEAT 339 348
FT REPEAT 349 368
FT REPEAT 359 368
FT CONFLICT 33 35
FT CONFLICT 89 89
FT CONFLICT 103 109
FT CONFLICT 141 148
FT CONFLICT 191 208
FT CONFLICT 241 241
FT CONFLICT 298 350
FT
FT
SQ SEQUENCE 394 AA: 43008 MM: 905CA71ECF4C87C2 CRC64;
Query Match 1.38; Score 7; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 416 GSSVTSL 422
Db 331 GSSVTSL 337
```

Search completed: May 12, 2003, 10:07:56  
Job time : 58 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 10:03:37 ; Search time 37 Seconds  
(without alignments)  
2923.641 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 525

Sequence: 1 MRANPKQTQAMPSETISLMKT.....KLMKRIAKILSLPIESLL 525

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size: 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database:

SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525	100.0	525	16 09JTT1	09JTT1 neisseria m
2	237	45.1	508	16 09JYU0	09JYU0 neisseria m
3	15	2.9	494	16 08Z028	08Z028 salmonella
4	15	2.9	495	16 08Z7M0	08Z7M0 salmonella
5	14	2.7	493	16 08X917	08X917 escherichia
6	14	2.7	514	16 09CKM5	09CKM5 pasteurilla
7	11	2.1	518	16 08UEX3	08UEX3 agrobacteri
8	11	2.1	521	16 09Z003	09Z003 rhizobium m
9	9	1.7	378	16 09X7V8	09X7V8 streptomyce
10	9	1.7	395	16 09KZ51	09KZ51 streptomyce
11	8	1.5	86	16 098A51	098A51 anabaena sp
12	8	1.5	161	16 08YX04	08YX04 anabaena sp
13	8	1.5	250	16 09RL45	09RL45 streptomyce
14	8	1.5	252	16 09CCW2	09CCW2 mycobacteri
15	8	1.5	272	17 08TYW8	08TYW8 methanopyru
16	8	1.5	466	16 098N23	098N23 rhizobium 1

17	8	1.5	466	16 0984B3	0984B3 rhizobium 1
18	8	1.5	652	16 09PEE1	09PEE1 xylella fas
19	8	1.5	662	16 09ZYV8	09ZYV8 rhizobium m
20	8	1.5	835	2 09KW59	09KW59 pectobacter
21	8	1.5	1347	10 093869	093869 arabidopsis
22	8	1.5	1854	10 09C606	09C606 arabidopsis
23	8	1.5	2599	10 09C727	09C727 arabidopsis
24	8	1.5	8243	5 096554	096554 cryptospori
25	7	1.3	16	11 09CTB2	09CTB2 mus musculu
26	7	1.3	71	16 P71622	P71622 mycobacteri
27	7	1.3	74	16 08RDE8	08RDE8 thermoaer
28	7	1.3	85	5 09VPC2	09VPC2 dirosophila
29	7	1.3	89	16 09A133	09A133 streptococ
30	7	1.3	94	10 09FRG7	09FRG7 oryza sativ
31	7	1.3	109	4 09BRH7	09BRH7 homo sapien
32	7	1.3	117	16 09K9X7	09K9X7 bacillus ha
33	7	1.3	123	2 09S324	09S324 paenibacill
34	7	1.3	124	10 042163	042163 arabidopsis
35	7	1.3	125	17 09YCM2	09YCM2 aeropyrum p
36	7	1.3	127	4 096MJ2	096MJ2 homo sapien
37	7	1.3	127	5 016513	016513 caenorhabdi
38	7	1.3	130	17 08TYE6	08TYE6 methanopyru
39	7	1.3	135	10 09LID1	09LID1 arabidopsis
40	7	1.3	136	16 026067	026067 helicobacte
41	7	1.3	136	16 09ZJ58	09ZJ58 helicobacte
42	7	1.3	141	13 09W619	09W619 cynops pyr
43	7	1.3	146	11 09DBD4	09DBD4 mus musculu
44	7	1.3	148	16 08UG70	08UG70 agrobacteri
45	7	1.3	157	16 08XGH2	08XGH2 bruceella me
46	7	1.3	158	17 08TT62	08TT62 methanosarc
47	7	1.3	159	17 09HQ30	09HQ30 halobacteri
48	7	1.3	162	5 015598	015598 entamoeba h
49	7	1.3	164	16 0983B0	0983B0 rhizobium 1
50	7	1.3	172	16 08ZNB4	08ZNB4 salmonella
51	7	1.3	172	16 08Z425	08Z425 salmonella
52	7	1.3	175	14 0991V8	0991V8 uncultured
53	7	1.3	178	5 044809	044809 caenorhabdi
54	7	1.3	178	12 098282	098282 moluscum c
55	7	1.3	183	16 0989U2	0989U2 rhizobium 1
56	7	1.3	184	12 091UK1	091UK1 red deer pa
57	7	1.3	188	16 08YEW7	08YEW7 bruceella me
58	7	1.3	199	5 09XX05	09XX05 caenorhabdi
59	7	1.3	202	17 08TUS0	08TUS0 methanosarc
60	7	1.3	206	10 08S6B0	08S6B0 oryza sativ
61	7	1.3	211	5 09UID3	09UID3 leishmania
62	7	1.3	218	10 039137	039137 arabidopsis
63	7	1.3	218	16 09PR77	09PR77 ureaplasma
64	7	1.3	221	2 09F4H4	09F4H4 yersinia en
65	7	1.3	221	16 08ZE34	08ZE34 yersinia pe
66	7	1.3	222	16 09KR04	09KR04 vibrio chol
67	7	1.3	222	16 08ZPE0	08ZPE0 salmonella
68	7	1.3	222	16 08Z7B3	08Z7B3 salmonella
69	7	1.3	223	16 09AA7	09AA7 caulobacter
70	7	1.3	224	7 030160	030160 homo sapien
71	7	1.3	225	5 08T1M6	08T1M6 dictyosteli
72	7	1.3	227	17 097AG9	097AG9 thermoplasm
73	7	1.3	229	16 09Z0C6	09Z0C6 rhizobium m
74	7	1.3	237	5 09W126	09W126 dirosophila
75	7	1.3	237	10 09F039	09F039 oryza sativ
76	7	1.3	239	10 09F0M6	09F0M6 arabidopsis
77	7	1.3	240	10 040456	040456 nicotiana t
78	7	1.3	243	17 097514	097514 sullolobus
79	7	1.3	243	16 097MW4	097MW4 clostridium
80	7	1.3	244	16 08UF88	08UF88 agrobacteri
81	7	1.3	246	10 09F0C5	09F0C5 zea mays (m
82	7	1.3	254	4 09BWV2	09BWV2 homo sapien
83	7	1.3	254	7 030631	030631 macaca mula
84	7	1.3	254	7 09TP70	09TP70 homo sapien
85	7	1.3	259	17 097CS9	097CS9 thermoplasm
86	7	1.3	260	5 09XVF7	09XVF7 caenorhabdi
87	7	1.3	260	10 09FEZ7	09FEZ7 glycine max
88	7	1.3	260	10 027620	027620 methanobact
89	7	1.3	262	10 09LJ97	09LJ97 arabidopsis













820	6	1.1	133	15	Q8UR90	human	immun	893	6	1.1	135	2	Q938W3	Q938W3 vibrio para
821	6	1.1	133	15	Q8UR89	human	immun	894	6	1.1	135	16	Q9JWP6	Q9JWP6 neisseria m
822	6	1.1	133	15	Q8UR88	human	immun	895	6	1.1	136	6	Q951A6	Q951A6 bos taurus
823	6	1.1	133	15	Q8UR87	human	immun	896	6	1.1	136	6	Q8WMA0	Q8WMA0 cynocephalu
824	6	1.1	133	15	Q8UR86	human	immun	897	6	1.1	136	11	Q9CQK5	Q9CQK5 mus musculu
825	6	1.1	133	15	Q8UR85	human	immun	898	6	1.1	136	16	Q8XXR2	Q8XXR2 ralsstonia s
826	6	1.1	133	15	Q8UR82	human	immun	899	6	1.1	136	17	Q9YAR0	Q9YAR0 aeryopyrum p
827	6	1.1	133	15	Q8UR81	human	immun	900	6	1.1	137	5	Q95OR7	Q95OR7 caenorhabd
828	6	1.1	133	15	Q8UR80	human	immun	901	6	1.1	137	10	Q8VXX9	Q8VXX9 arabidopsis
829	6	1.1	133	15	Q8UR79	human	immun	902	6	1.1	137	10	Q9ZT46	Q9ZT46 cucurbita m
830	6	1.1	133	15	Q8UR74	human	immun	903	6	1.1	137	16	Q9PAB3	Q9PAB3 xyliella fas
831	6	1.1	133	15	Q8UR73	human	immun	904	6	1.1	137	17	Q978M7	Q978M7 thermoplasm
832	6	1.1	133	15	Q8UR72	human	immun	905	6	1.1	138	2	Q53074	Q53074 lactococcus
833	6	1.1	133	15	Q8UR71	human	immun	906	6	1.1	138	4	Q9UML6	Q9UML6 homo sapien
834	6	1.1	133	15	Q8UR66	human	immun	907	6	1.1	138	9	Q9AF62	Q9AF62 streptococc
835	6	1.1	133	15	Q8UR65	human	immun	908	6	1.1	138	12	Q68219	Q68219 hepatitis c
836	6	1.1	133	15	Q8UR64	human	immun	909	6	1.1	138	12	Q68220	Q68220 hepatitis c
837	6	1.1	133	15	Q8UR61	human	immun	910	6	1.1	139	2	Q50234	Q50234 zymomonas m
838	6	1.1	133	15	Q8UR60	human	immun	911	6	1.1	139	2	Q9RF30	Q9RF30 leptolyngby
839	6	1.1	133	15	Q8UR59	human	immun	912	6	1.1	139	10	Q8W3Z5	Q8W3Z5 nicotiana t
840	6	1.1	133	15	Q8UR58	human	immun	913	6	1.1	139	11	Q9D6D7	Q9D6D7 mus musculu
841	6	1.1	133	15	Q8UR57	human	immun	914	6	1.1	139	12	Q9J1L6	Q9J1L6 newcastle d
842	6	1.1	133	15	Q8UR55	human	immun	915	6	1.1	139	16	Q8RYX1	Q8RYX1 brucella me
843	6	1.1	133	15	Q8UR52	human	immun	916	6	1.1	140	2	Q939G2	Q939G2 pseudomonas
844	6	1.1	133	15	Q8UR51	human	immun	917	6	1.1	140	11	Q9EOP4	Q9EOP4 mus musculu
845	6	1.1	133	15	Q8UR50	human	immun	918	6	1.1	140	12	Q96636	Q96636 bovine herp
846	6	1.1	133	15	Q8UR48	human	immun	919	6	1.1	141	2	Q9R215	Q9R215 pseudomonas
847	6	1.1	133	15	Q8UR47	human	immun	920	6	1.1	141	17	Q26561	Q26561 methanobact
848	6	1.1	133	15	Q8UR46	human	immun	921	6	1.1	142	2	Q50186	Q50186 mycoplasma
849	6	1.1	133	15	Q8UR44	human	immun	922	6	1.1	142	4	Q9BF74	Q9BF74 homo sapien
850	6	1.1	133	15	Q8UR43	human	immun	923	6	1.1	142	16	Q91125	Q91125 pseudomonas
851	6	1.1	133	15	Q8UR41	human	immun	924	6	1.1	142	16	Q05787	Q05787 mycobacteri
852	6	1.1	133	15	Q8UR40	human	immun	925	6	1.1	142	17	Q58906	Q58906 pyrococcus
853	6	1.1	133	15	Q8UR39	human	immun	926	6	1.1	143	5	Q76748	Q76748 stentor coe
854	6	1.1	133	15	Q8UR34	human	immun	927	6	1.1	143	10	Q8RUZ6	Q8RUZ6 arabidopsis
855	6	1.1	133	15	Q8UR33	human	immun	928	6	1.1	143	12	Q99CZ3	Q99CZ3 bovine herp
856	6	1.1	133	15	Q8UR32	human	immun	929	6	1.1	144	2	Q93A89	Q93A89 nitrososplir
857	6	1.1	133	15	Q8UR31	human	immun	930	6	1.1	144	2	Q939G3	Q939G3 pseudomonas
858	6	1.1	133	15	Q8UR30	human	immun	931	6	1.1	144	6	Q9RU38	Q9RU38 oryctolagus
859	6	1.1	133	15	Q8UR29	human	immun	932	6	1.1	144	11	Q9JYW0	Q9JYW0 mus musculu
860	6	1.1	133	15	Q8UR28	human	immun	933	6	1.1	144	12	Q9PYW7	Q9PYW7 xestia c-ni
861	6	1.1	133	15	Q8UR27	human	immun	934	6	1.1	144	16	Q54184	Q54184 streptomyc
862	6	1.1	133	15	Q8UR26	human	immun	935	6	1.1	144	17	Q8TWJ2	Q8TWJ2 methanopyru
863	6	1.1	133	15	Q8UR25	human	immun	936	6	1.1	145	4	Q9H0Z2	Q9H0Z2 homo sapien
864	6	1.1	133	15	Q8UR24	human	immun	937	6	1.1	145	5	Q9N3V4	Q9N3V4 caenorhabd
865	6	1.1	133	15	Q8UR23	human	immun	938	6	1.1	145	5	Q9NMB6	Q9NMB6 leishmania
866	6	1.1	133	15	Q8UR22	human	immun	939	6	1.1	145	5	Q9N864	Q9N864 leishmania
867	6	1.1	133	15	Q8UR21	human	immun	940	6	1.1	145	12	Q6S9404	Q6S9404 pseudorabie
868	6	1.1	133	15	Q8UR19	human	immun	941	6	1.1	145	12	Q9YPT0	Q9YPT0 ipomoea yel
869	6	1.1	133	15	Q8UR18	human	immun	942	6	1.1	145	16	Q9CEK4	Q9CEK4 lactococcus
870	6	1.1	133	15	Q8UR17	human	immun	943	6	1.1	145	16	Q97T28	Q97T28 streptococc
871	6	1.1	133	15	Q8UR16	human	immun	944	6	1.1	145	16	Q8ZJW5	Q8ZJW5 salmonella
872	6	1.1	133	15	Q8UR15	human	immun	945	6	1.1	145	16	Q8Z0V1	Q8Z0V1 salmonella
873	6	1.1	133	15	Q8UR13	human	immun	946	6	1.1	146	2	Q9Y414	Q9Y414 escherichia
874	6	1.1	133	15	Q8UR08	human	immun	947	6	1.1	146	2	Q45052	Q45052 borrelia bu
875	6	1.1	133	15	Q8UR07	human	immun	948	6	1.1	146	9	Q9JMN7	Q9JMN7 bacterioph
876	6	1.1	133	15	Q8UR02	human	immun	949	6	1.1	146	12	Q41076	Q41076 paramecium
877	6	1.1	133	15	Q8UR01	human	immun	950	6	1.1	146	16	Q51615	Q51615 borrelia bu
878	6	1.1	133	15	Q8UQZ9	human	immun	951	6	1.1	147	2	Q936Z5	Q936Z5 pseudomonas
879	6	1.1	133	15	Q8UQZ2	human	immun	952	6	1.1	147	16	Q26086	Q26086 helicobacte
880	6	1.1	133	15	Q8UQZ1	human	immun	953	6	1.1	147	16	Q8XT00	Q8XT00 ralsstonia s
881	6	1.1	133	15	Q8UQZ0	human	immun	954	6	1.1	148	6	Q29234	Q29234 sus scrofa
882	6	1.1	133	15	Q8UQY9	human	immun	955	6	1.1	148	10	Q49305	Q49305 arabidopsis
883	6	1.1	133	15	Q8UQY7	human	immun	956	6	1.1	148	12	Q9Q8B7	Q9Q8B7 yaba monkey
884	6	1.1	133	15	Q8UQY6	human	immun	957	6	1.1	148	16	Q9AB22	Q9AB22 calobacter
885	6	1.1	133	15	Q8UQY5	human	immun	958	6	1.1	148	16	Q97PX5	Q97PX5 streptococc
886	6	1.1	133	15	Q8UQY4	human	immun	959	6	1.1	148	17	Q8U067	Q8U067 pyrococcus
887	6	1.1	133	15	Q8UQX6	human	immun	960	6	1.1	149	2	Q56748	Q56748 wolfinella s
888	6	1.1	133	15	Q8UQW4	human	immun	961	6	1.1	149	2	Q59588	Q59588 mycobacteri
889	6	1.1	133	15	Q8UQW3	human	immun	962	6	1.1	149	10	Q9ZT47	Q9ZT47 cucurbita m
890	6	1.1	133	15	Q8UQW1	human	immun	963	6	1.1	149	11	Q8R4G5	Q8R4G5 mus musculu
891	6	1.1	133	15	Q8YQJ8	brucella me	immun	964	6	1.1	149	12	Q12671	Q12671 colobus mon
892	6	1.1	134	13	Q73933	gallus gall	965	6	1.1	149	16	Q98L56	Q98L56 rhizobium l	

966	6	1.1	149	16	08X14	08x14 clostridium
967	6	1.1	149	16	09L14	09l14 streptomyce
968	6	1.1	150	2	09R16	09r16 comamonas t
969	6	1.1	150	5	023703	023703 ctenolepism
970	6	1.1	150	10	022550	022550 triticum ae
971	6	1.1	150	10	09LNG6	09lng6 arabidopsis
972	6	1.1	150	17	09YD88	09ydb8 aeropyrum p
973	6	1.1	151	2	068739	068739 yersinia pe
974	6	1.1	151	3	09HG48	09hg48 pneumocyst
975	6	1.1	151	5	09NB67	09nb67 drosophila
976	6	1.1	151	5	09VC44	09vc44 drosophila
977	6	1.1	151	11	09CZC0	09czc0 mus musculu
978	6	1.1	152	2	08VOP3	08vop3 burkholderi
979	6	1.1	152	4	09HA18	09ha18 homo sapien
980	6	1.1	152	12	09DHL8	09dhl8 yaba-like d
981	6	1.1	152	16	092UX9	092ux9 rhizobium m
982	6	1.1	153	16	08XXR7	08xxr7 raltostonia s
983	6	1.1	154	16	09CH33	09ch33 lactococcus
984	6	1.1	155	10	09C886	09c886 arabidopsis
985	6	1.1	155	15	096774	096774 feline immu
986	6	1.1	155	16	08XFK3	08xfk3 raltostonia s
987	6	1.1	156	6	08NNA7	08nna7 actinobac
988	6	1.1	156	10	08WNA6	08wna6 actus trivi
989	6	1.1	156	10	09FTJ1	09ftj1 oryza sativ
990	6	1.1	157	5	018506	018506 strongyloid
991	6	1.1	157	5	08T013	08t013 drosophila
992	6	1.1	157	6	09TUT6	09tut6 cryctolagus
993	6	1.1	157	12	08WJ67	08wj67 jamesonia l
994	6	1.1	157	12	08VAG6	08vag6 white spot
995	6	1.1	157	16	08XJ05	08xj05 clostridium
996	6	1.1	158	2	09F9J1	09f9j1 comamonas s
997	6	1.1	158	4	09UKP8	09ukp8 homo sapien
998	6	1.1	158	5	017953	017953 caenorhabdi
999	6	1.1	158	11	09CUN4	09cun4 mus musculu
1000	6	1.1	158	15	096778	096778 feline immu

## ALIGNMENTS

RESULT 1

ID	09JUT1	PRELIMINARY:	PRT:	525 AA.
AC	09JUT1;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Phospholipase D-family protein.			
GN	NMA1646.			
OS	Neisseria meningitidis (serogroup A).			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=65699;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=2491 / SEROGROUP A / SEROTYPE 4A;			
RX	MEDLINE=20222556; PubMed=10761919;			
RA	Patkhili J., Achman M., James K.D., Bentley S.D., Churcher C.,			
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,			
RA	Jagers R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,			
RA	Rajandream M.A., Rutherford K.M., Mungall K., Quail M.A.,			
RA	Whitehead S., Spratt B.G., Barrell B.G.,			
RT	"Complete DNA sequence of a serogroup A strain of Neisseria			
RL	meningitidis 22491."			
DR	EMBL; AL162756; CAB84874.1; -			
DR	InterPro; IPR001736; PLD.			
DR	Pfam; PF00614; PLDC; 2.			
DR	SMART; SM00155; PLDC; 2.			
KW	Complete proteome.			
SEQUENCE	525 AA; 59305 MM; ICFB/AC582F1B02 CRC64;			

Query Match 100.0%; Score 525; DB 16; Length 525;

Best Local Similarity 100.0%; Pred. No. 0; Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRANPKTOAMPSETTISLMKTRSLISLCLLCSCSSMLPLEERTSRFPNTSKPRLDN 60

DB 1 MRANPKTOAMPSETTISLMKTRSLISLCLLCSCSSMLPLEERTSRFPNTSKPRLDN 60

QY 61 IIOIRHTPHTNGSLIDYLLNDPHEAFARAAALIESAHSLSLQYIWRNDISGRLLFNLY 120

DB 61 IIOIRHTPHTNGSLIDYLLNDPHEAFARAAALIESAHSLSLQYIWRNDISGRLLFNLY 120

QY 121 YLAEGSVRRLLDNNRTGDLDDLALDSHNTVEVRLFPNPKRMALGYLDFPRLL 180

DB 121 YLAEGSVRRLLDNNRTGDLDDLALDSHNTVEVRLFPNPKRMALGYLDFPRLL 180

QY 121 YLAEGSVRRLLDNNRTGDLDDLALDSHNTVEVRLFPNPKRMALGYLDFPRLL 180

DB 121 YLAEGSVRRLLDNNRTGDLDDLALDSHNTVEVRLFPNPKRMALGYLDFPRLL 180

QY 181 NRRMKNKSTFADNRATILIGRNIGDEYFVGEYEDYFADDIATGSGVGEVSHDEPRYWA 240

DB 181 NRRMKNKSTFADNRATILIGRNIGDEYFVGEYEDYFADDIATGSGVGEVSHDEPRYWA 240

QY 241 SSHAHNATRIISGNIGKGLQALGYNDETSRHALLRYRETVEQSPLYOKIQGRIDMOSV 300

DB 241 SSHAHNATRIISGNIGKGLQALGYNDETSRHALLRYRETVEQSPLYOKIQGRIDMOSV 300

QY 301 QPRLISDDPAKGLDRRRKPIAGRLQDALKQPKSVLYSPFVPTKGTALAKLYOD 360

DB 301 QPRLISDDPAKGLDRRRKPIAGRLQDALKQPKSVLYSPFVPTKGTALAKLYOD 360

QY 361 GIDVTVLINSLOATDVAAVHSGVYKRPDLKAGIKLYELOPNHNAVPAATKDKLTGSSVT 420

DB 361 GIDVTVLINSLOATDVAAVHSGVYKRPDLKAGIKLYELOPNHNAVPAATKDKLTGSSVT 420

QY 421 SLHAKFTYDGRKRFITGSPNLDPRSARLTEMGVYIESPKINAEOMERTLADTSPYARV 480

DB 421 SLHAKFTYDGRKRFITGSPNLDPRSARLTEMGVYIESPKINAEOMERTLADTSPYARV 480

QY 481 TLDNRHRLQWHPATRTKTPNEPEAKIMKRIAKILSLPIESL 525

DB 481 TLDNRHRLQWHPATRTKTPNEPEAKIMKRIAKILSLPIESL 525

RESULT 2

ID	09JYU0	PRELIMINARY:	PRT:	508 AA.
AC	09JYU0;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Cardiolipin synthetase family protein.			
GN	NMB1434.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxID=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / SEROGROUP B;			
RX	MEDLINE=20175755; PubMed=10710307;			
RA	Tetzelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,			
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,			
RA	Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,			
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,			
RA	Mason T., Clecko A., Parksey D.S., Blair E., Clifton E., Clark E.B.,			
RA	Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,			
RA	Gill J., Scarlato V., Masiugani V., Piza M., Grandi G., Sun L.,			
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.,			
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain			
RL	MC58."			
DR	EMBL; AE002493; AAF11795.1; -			
DR	TIGR; NMB1434; -			
DR	InterPro; IPR001736; PLD.			
DR	Pfam; PF00614; PLDC; 2.			
DR	SMART; SM00155; PLDC; 2.			
KW	Complete proteome.			

SQ SEQUENCE 508 AA; 57350 MW; A2DE6CAC47CA25D5 CRC64;  
 Query Match 45.1%; Score 237; DB 16; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 8.2e-234;  
 Matches 237; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 MKTRSLISLCLLSCSSWMLPPEBERTSRHNTSKPYRLNLIQIRHTPHNGISDIY 77  
 DB 1 MKTRSLISLCLLSCSSWMLPPEBERTSRHNTSKPYRLNLIQIRHTPHNGISDIY 60  
 QY 78 LNDPHEAFAARALIESAESHSDLOXYTRNDISGRLEFNLYLAERGVRRLDDN 137  
 DB 61 LNDPHEAFAARALIESAESHSDLOXYTRNDISGRLEFNLYLAERGVRRLDDN 120  
 QY 138 NTRGDLDDLALDSHPNIEVRLFPVLRKRALGYLTDFPRLRMHNSFTADNRRT 197  
 DB 121 NTRGDLDDLALDSHPNIEVRLFPVLRKRALGYLTDFPRLRMHNSFTADNRRT 180  
 QY 198 LGGRTIGDEYFKVGEDTFVADDIATGSGVGEVSHDFDRYMAASHANATRIIRSG 254  
 DB 181 LGGRTIGDEYFKVGEDTFVADDIATGSGVGEVSHDFDRYMAASHANATRIIRSG 237

RESULT 3  
 ID Q8ZQ28 PRELIMINARY; PRT; 494 AA.  
 AC Q8ZQ28;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 GN Putative phospholipase.  
 OS YMDC OR STM1148.  
 OC Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OK NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;  
 RA MEDLINE-21534948; PubMed-11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 DR EMBL: AE008749; AAL20078.1; -;  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 494 AA; 56217 MW; 571C01BFC4FCB288 CRC64;

Query Match 2.9%; Score 15; DB 16; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 LDSHPNIEVRLFPNF 163  
 DB 115 LDSHPNIEVRLFPNF 129

RESULT 4  
 ID Q8Z7M0 PRELIMINARY; PRT; 495 AA.  
 AC Q8Z7M0;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Hypothetical protein STY1185.  
 GN STY1185.

OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OK NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CT18;  
 RX MEDLINE-21534947; PubMed-11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsels K.,  
 RA Krogli A., Larsen T.S., Leather S., Moulle S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrett B.G.;  
 RT "Complete genome sequence of a multiple drug resistant *Salmonella*  
 enterica serovar typhi CT18.";  
 RL Nature 413:848-852(2001).  
 DR EMBL: AL627269; CAD08272.1; -;  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 495 AA; 56348 MW; 4555257D8DD83AB9 CRC64;

Query Match 2.9%; Score 15; DB 16; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 LDSHPNIEVRLFPNF 163  
 DB 116 LDSHPNIEVRLFPNF 130

RESULT 5  
 ID Q8X917 PRELIMINARY; PRT; 493 AA.  
 AC Q8X917;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 GN Putative synthase.  
 DE YMDC OR Z1680 OR ECS1424.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OK NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE-21074935; PubMed-11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobbeck E.J., Davis N.W., Lim A., Dialanta E.T., Potomousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-O157:H7 / RIMD 0509952;  
 RX MEDLINE-21156231; PubMed-11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*"  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL: AE005315; AAG55792.1; -;  
 DR EMBL: AP002555; BAB34847.1; -;

DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Complete proteome.  
 SO SEQUENCE 493 AA; 55929 MW; DEC407F83D7D1CFF CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 14; DB 16; Length 493;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDDNT 139  
 DB 91 RGVRRLLDDNT 104

RESULT 6  
 O9CKM5 PRELIMINARY; PRT; 514 AA.  
 AC O9CKM5:  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical protein PM1586.  
 GN PM1586.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.,  
 RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL: AE006196; AAK03670.1; -  
 DR InterPro: IPR01736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 514 AA; 57882 MW; C65B0EB65D80F216 CRC64;  
 SO SEQUENCE 514 AA; 57882 MW; C65B0EB65D80F216 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 14; DB 16; Length 514;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 179 RLRMRHNSFTAD 192  
 DB 170 RLRMRHNSFTAD 183

RESULT 7  
 O8UEX3 PRELIMINARY; PRT; 518 AA.  
 AC O8UEX3:  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Phospholipase D family protein.  
 GN ATU1630 OR AGR\_C\_3007.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D., Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Seephimmachak C., Wu Z., Romero P., Gordon D.,

RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58.";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askew M., Halling C., Mullin L.,  
 RA Hummel K., Gordon J., Vaudin M., Tarchoux O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL: AE009120; AAL42632.1; -  
 DR EMBL: AE008085; AAK87408.1; -  
 KW Complete proteome.  
 SO SEQUENCE 518 AA; 58095 MW; F08804A62B810AE2 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 11; DB 16; Length 518;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDD 136  
 DB 115 RGVRRLLDD 125

RESULT 8  
 O920Q3 PRELIMINARY; PRT; 521 AA.  
 ID O920Q3:  
 AC O920Q3:  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical transmembrane protein SMC04448.  
 GN R01254 OR SMC04448.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,  
 RA Pohl T., Portetelle D., Puehler A., Purrelle B., Rampeger U.,  
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galbette F.,  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 DR EMBL: AL591786; CAC45833.1; -  
 DR InterPro: IPR01736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 521 AA; 57974 MW; 4CD7CD8D6AE2E766 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 11; DB 16; Length 521;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVRRLLDD 136  
 DB 116 RGVRRLLDD 126

```

RESULT 9
ID 09X7V8 PRELIMINARY: PRT: 378 AA.
AC 09X7V8:
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative polyprenyl synthase.
GN SC06763 OR SC6A5.12.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RL Oliver K., Harris D.;
RT Submitted (Mar-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RT Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (Mar-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinshl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL049485; CAB39696.1; -
DR InterPro: IPR000794; Ketoacyl_synth.
DR InterPro: IPR000092; Polyprenyl_synth.
DR Pfam: PF00348; polyprenyl_synth. 1.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; UNKNOWN.1.
DR PROSITE: PS00723; POLYPRENYL_SYNTHET.1; UNKNOWN.1.
SQ SEQUENCE 378 AA; 4011 MW; B16787B624477F3C CRC64;

Query Match 1.7%; Score 9; DB 16; Length 378;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 FAARALIE 94
DB 322 FAARALIE 330

```

```

OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Brown S.P., Harris D.;
RT Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinshl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL354616; CAB89753.1; -
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDC; 2.
DR SMART: SM00155; PLDC; 1.
SQ SEQUENCE 395 AA; 44541 MW; 7D2021762EBB697 CRC64;

Query Match 1.7%; Score 9; DB 16; Length 395;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 GYVRRLLD 135
DB 87 GYVRRLLD 95

RESULT 11
ID 098A51 PRELIMINARY: PRT: 86 AA.
AC 098A51:
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE Transposase.
GN MSL6147.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MAEF30309;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsuno M., Matsuno A.,

```

RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003008: BAB52486.1;  
 KW Complete proteome.  
 SQ SEQUENCE 86 AA; 9693 MW; FF269E9BBD20CAA CRC64;

Query Match 1.5%; Score 8; DB 16; Length 86;  
 Best Local Similarity 100.0%; Pred. No. 8.4;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 LAERGV 129  
 DB 24 LAERGV 31

RESULT 12

OYX04 PRELIMINARY; PRT; 161 AA.  
 AC OYX04;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein AL1158.  
 GN AL1158.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Kishida Y., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL: AP003584: BAB73115.1;  
 DR InterPro: IPR000999: RNase\_3.  
 DR SMART: SM00535: RIBOC\_1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 161 AA; 18438 MW; C9D6A5FC8421CA76 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 173 YLTDPR 180  
 DB 141 YLTDPR 148

RESULT 13

OYRL45 PRELIMINARY; PRT; 250 AA.  
 ID OYRL45;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein SC00441.  
 GN SC00441 OR SCF51A.19.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Croft A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL121596: CAB56671.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 250 AA; 25510 MW; 6F115EB480AD1700 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 250;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 LAERGV 129  
 DB 50 LAERGV 57

RESULT 14

OYCCW2 PRELIMINARY; PRT; 252 AA.  
 AC OYCCW2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE Hypothetical protein ML0333.  
 GN ML0333.  
 OS Mycobacterium leprae.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21126732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus";  
 RT Nature 409:1007-1011(2001).  
 DR EMBL: AL583918: CAC29841.1;  
 DR Leprona: ML0333;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 252 AA; 26453 MW; 98FB2B21AF8592C CRC64;

Query Match 1.5%; Score 8; DB 16; Length 252;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 LAERGV 129  
 DB 52 LAERGV 59

RESULT 15

OYTW8 PRELIMINARY; PRT; 272 AA.  
 ID OYTW8;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Imidazoleglycerol-phosphate synthase.  
 GN HISF OR MK0173.  
 OS Methanopyrus kandleri.  
 OC Archaea: Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 RX NCBI\_TaxId=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Mal'kh A.G., Koonin E.V., Kozaykin S.A.;  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL: AE010316; AAM01390.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 272 AA; 29848 MW; CA7F9018E3A1A0CC CRC64;

Query Match 1.5%; Score 8; DB 16; Length 272;  
 Best Local Similarity 100.0%; Pred. No. 23;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 124 AERGVAVR 131  
 |||||  
 Db 263 AERGVAVR 270

RESULT 16  
 Q98NZ3 PRELIMINARY; PRT; 466 AA.  
 AC Q98NZ3;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Phospholipase D-family protein.  
 GN MUR9675.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Plasmid PMB.  
 OC Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 RX NCBI\_TaxId=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003017; BAB54862.1; -.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Plasmid; Complete proteome.  
 SQ SEQUENCE 466 AA; 51793 MW; 4693C68F20F94F3F CRC64;

Query Match 1.5%; Score 8; DB 16; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 38;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 373 ATDVAAVH 380  
 |||||  
 Db 315 ATDVAAVH 322

RESULT 17

O984B3 PRELIMINARY; PRT; 466 AA.  
 AC O984B3;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Cardiolipin synthetase family protein.  
 GN MUR8077.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 RX NCBI\_TaxId=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003013; BAB53717.1; -.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 466 AA; 52117 MW; 693301E508A0A788 CRC64;

Query Match 1.5%; Score 8; DB 16; Length 466;  
 Best Local Similarity 100.0%; Pred. No. 38;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 373 ATDVAAVH 380  
 |||||  
 Db 315 ATDVAAVH 322

RESULT 18  
 O9PEE1 PRELIMINARY; PRT; 652 AA.  
 AC O9PEE1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Cardiolipin synthase.  
 GN XF1087.  
 OS Xylella fastidiosa.  
 OC Bacteria: Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xylella.  
 RX NCBI\_TaxId=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 RA Krieger J.E., Kuramie E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miraca E.C., Miyaki C.Y., Montello-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quadrigio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*.";  
 RL Nature 406:151-159(2000).  
 DR EMBL: AB003945; AAF83897.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 652 AA; 72962 MW; 709E6A18A348D78C CRC64;

Query Match  
 Best Local Similarity 1.5%; Score 8; DB 16; Length 652;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RGVVRLL 133  
 DB 126 RGVVRLL 133

RESULT 19  
 ID 092VJ8 PRELIMINARY; PRT: 662 AA.  
 AC 092VJ8;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE putative methylcrotonoyl-CoA carboxylase biotinylated subunit  
 DE protein (EC 6.4.1.4).  
 GN MCB OR RB0703 OR SMB21124.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Plasmid pSymb (megaplasmid 2).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OX NCBI\_TaxID=382;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,  
 RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,  
 RA Golding B., Puehler A.;  
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont *Sinorhizobium meliloti*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL: AL603644; CAC49103.1; -  
 DR InterPro: IPR001882; Biotin\_attach.  
 DR InterPro: IPR000089; Biotin\_1ipoyl.  
 DR InterPro: IPR000901; CPSase  
 DR Pfam: PF02785; Biotin\_carb\_C; 1.  
 DR Pfam: PF00364; Biotin\_1ipoyl; 1.  
 DR Pfam: PF00289; CPSase\_L\_chain; 1.  
 DR Pfam: PF02786; CPSase\_L\_D2; 1.  
 DR PROSITE: PS00188; BIOTIN; UNKNOWN.1.  
 DR PROSITE: PS00867; CPSASE\_2; UNKNOWN.1.  
 KW ligase; plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 662 AA; 71317 MW; 3C5C9B1421FC1B5B CRC64;

Query Match  
 Best Local Similarity 1.5%; Score 8; DB 16; Length 662;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 20  
 ID 09KM59 PRELIMINARY; PRT: 835 AA.  
 AC 09KM59;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Tail protein.  
 OS Pectobacterium carotovorum subsp. carotovorum.  
 OC Pectobacterium; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=555;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CGE234M403;  
 RA Hirota M., Nguyen A.H., Sano T., Kaneko J., Kamio Y.;  
 RT "Nucleotide sequence of bacteriocin from *Erwinia carotovora*.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB045036; BAA96858.1; -  
 SQ SEQUENCE 835 AA; 90498 MW; 7E2A24F5D074714F CRC64;

Query Match  
 Best Local Similarity 1.5%; Score 8; DB 2; Length 835;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 372 QATDVAAV 379  
 DB 190 QATDVAAV 197

RESULT 21  
 ID 038969 PRELIMINARY; PRT: 1347 AA.  
 AC 038969;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical 153.9 kDa protein.  
 GN SABRE.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WS ECTOTYPE;  
 RX MEDLINE=95172383; PubMed=7867930;  
 RA Aeschbacher R.A., Hauser M.T., Feldmann K.A., Benfey P.N.;  
 RT "The SABRE gene is required for normal cell expansion in  
 RT Arabidopsis.";  
 RL Genes Dev. 9:330-340(1995).  
 DR EMBL: U19134; AAC49734.1; JOINED.  
 DR EMBL: U19121; AAC49734.1; JOINED.  
 DR EMBL: U19123; AAC49734.1; JOINED.  
 DR EMBL: U19122; AAC49734.1; JOINED.  
 DR EMBL: U19124; AAC49734.1; JOINED.  
 DR EMBL: U19125; AAC49734.1; JOINED.  
 DR EMBL: U19126; AAC49734.1; JOINED.  
 DR EMBL: U19127; AAC49734.1; JOINED.  
 DR EMBL: U19128; AAC49734.1; JOINED.  
 DR EMBL: U19129; AAC49734.1; JOINED.  
 DR EMBL: U19130; AAC49734.1; JOINED.  
 DR EMBL: U19131; AAC49734.1; JOINED.  
 DR EMBL: U19132; AAC49734.1; JOINED.  
 DR EMBL: U19133; AAC49734.1; JOINED.  
 DR EMBL: U19133; AAC49734.1; JOINED.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1347 AA; 153905 MW; 0011A29D5D8DDDD0 CRC64;

Query Match  
 Best Local Similarity 1.5%; Score 8; DB 10; Length 1347;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 324 GRLQDALK 331  
 DB 405 GRLQDALK 412



Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SDDPAKGL 313  
 |||||  
 Db 264 SDDPAKGL 271

RESULT 22  
 09C606 PRELIMINARY; PRT; 1854 AA.

AC 09C606; PRT; 1854 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE Hypothetical 210.0 kDa protein (Fragment).

GN T18124.18.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maitl R., Marzilli A., Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utebback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."

RL Nature 408:816-820(2000).

DR EMBL: AC079131; AAG50770.1; -.

KW Hypothetical protein.

FT NON\_TER 1 1

SQ SEQUENCE 1854 AA; 209958 MW; C2A5PB46A994C94C CRC64;

Query Match 1.5%; Score 8; DB 10; Length 1854;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SDDPAKGL 313  
 |||||  
 Db 769 SDDPAKGL 776

RESULT 23  
 09C727 PRELIMINARY; PRT; 2599 AA.

AC 09C727; PRT; 2599 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE Hypothetical 292.4 kDa protein.

GN P16M22.5.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CV. COLUMBIA;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S., White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y., Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W., Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K., Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y., Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L., Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E., Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B., Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P., Lin X., Liu S.X., Liu Z.A., Luros J.S., Maitl R., Marzilli A., Miltscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I., Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D., Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M., Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D., Utebback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M., Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana."

RL Nature 408:816-820(2000).

DR EMBL: AC073943; AAG50951.1; -.

KW Hypothetical protein.

FT NON\_TER 2599 AA; 292445 MW; 6FAF3CCA5924A3C CRC64;

SQ SEQUENCE 2599 AA; 292445 MW; 6FAF3CCA5924A3C CRC64;

Query Match 1.5%; Score 8; DB 10; Length 2599;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 SDDPAKGL 313  
 |||||  
 Db 1516 SDDPAKGL 1523

RESULT 24  
 096554 PRELIMINARY; PRT; 8243 AA.

AC 096554; PRT; 8243 AA.

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Type I fatty acid synthase.

GN FAS1.

OS Cryptosporidium parvum.

OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;

OC Cryptosporidiidae; Cryptosporidium.

OX NCBI\_TaxID=5807;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=KSU-1;

RX MEDLINE=20155414; PubMed=10693747;

RA Zhu G., Marchewka M.J., Woods K.M., Upton S.J., Keithly J.S.

RT "Molecular analysis of a Type I fatty acid synthase in Cryptosporidium parvum."

RL Mol. Biochem. Parasitol. 105:253-260(2000).

DR EMBL: AF082993; AAC99407.1; -.

DR HSSP: P28304; 100R.

DR InterPro: IPR001227; AC transferase.

DR InterPro: IPR002085; Adh\_zn\_family.

DR InterPro: IPR000873; AMP-bind.

DR InterPro: IPR000794; ketoacyl-synt.

DR InterPro: IPR003880; Pantane\_attach.

DR Pfam: PF00698; Acyl\_transf. 3.

DR Pfam: PF00107; adh\_zinc. 3.

DR Pfam: PF00501; AMP-binding; 2.

DR Pfam: PF00109; ketoacyl-synt. 3.

DR Pfam: PF02801; ketoacyl-synt\_C; 3.

DR Pfam: PF00507; pp-binding; 4.

DR PROSITE: PS50075; ACP\_DOMAIN; 4.

DR PROSITE: PS00455; AMP\_BINDING; UNKNOWN\_1.

DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; UNKNOWN\_2.

DR PROSITE: PS00012; PHOSPHOPANTETHEINE; UNKNOWN\_2.

KW Phosphopantetheine.  
 SQ SEQUENCE 8243 AA; 920806 MW; F5A1BC8E4B3E37C CRC64;  
 Query Match 1.5%; Score 8; DB 5; Length 8243;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 GLTGSSTV 420  
 DB 1045 GLTGSSTV 1052

RESULT 25  
 Q9CTB2 PRELIMINARY; PRT; 16 AA.  
 AC 09CTB2;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Nitrogen fixation gene 1 (S. cerevisiae) (Fragment).  
 GN NFS1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana T.,  
 Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,  
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 Ra Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Mashio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.O., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL: AK004077; BAB23157.1;  
 DR MGI:1316706; NFS1.  
 FT NON\_TER 1  
 SQ SEQUENCE 16 AA; 1881 MW; 029ECFC9BC205263 CRC64;

Query Match 1.3%; Score 7; DB 11; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 LVQDGD 363  
 DB 1 LVQDGD 7

RESULT 26  
 P71622 PRELIMINARY; PRT; 71 AA.  
 AC P71622;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein RV2830c.  
 GN RV2830C OR MTCY16B7.12.  
 OS Mycobacterium tuberculosis.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Horsby T., Jagsels K., Krogh A., McLaren J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skellern S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544(1998).  
 DR EMBL: Z81331; CAB03656.1;  
 DR Tuberculosis; RV2830c;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 71 AA; 7483 MW; 7F79A4D85601879A CRC64;

Query Match 1.3%; Score 7; DB 16; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 75;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 AKILSL 519  
 DB 8 AKILSL 14

RESULT 27  
 Q8RDE8 PRELIMINARY; PRT; 74 AA.  
 AC Q8RDE8;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Sigma-F transcribed protein csfB.  
 GN TTE0092.  
 OS Thermoaerobacter tengcongensis.  
 OC Bacteria; Firmicutes; Bacilli; Clostridium group; Clostridia;  
 OC Thermoaerobacteriales; Thermoaerobacteriaceae; Thermoaerobacter.  
 OX NCBI\_TaxID=119072;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB47 / JCM11007;  
 RX MEDLINE=21992816; PubMed=11997336;  
 RA Bao O., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
 RT "A complete sequence of T. tengcongensis genome."  
 RL Genome Res. 12:689-700(2002).  
 DR EMBL: AE012983; AAM23396.1;  
 KW Complete proteome.  
 SQ SEQUENCE 74 AA; 9014 MW; 060CB42623456B60 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 VQDGDV 364  
 DB 21 VQDGDV 27

RESULT 28  
 Q9VPC2 PRELIMINARY; PRT; 85 AA.  
 AC Q9VPC2;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TREMBlrel. 13, last annotation update)  
 DE CG4186 protein.  
 OS CG4186.  
 GN Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mettel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sutskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Weissman D.A., Weissbach G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003592; AAF51633.1;  
 DR FlyBase: FBgn0040634; CG4186.  
 SQ SEQUENCE 85 AA; 10104 MW; 5547E24F36573CFD CRC64;  
  
 Query Match 1.3%; Score 7; DB 5; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 38 LPPLEER 44  
 DB 64 LPPLEER 70  
  
 RESULT 29  
 ID 09A133 PRELIMINARY; PRT; 89 AA.  
 AC 09A133;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Hypothetical protein SPY0492.  
 GN SPY0492.  
 OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 NCBI\_TaxID=1314;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;  
 RX MEDLINE=21192684; PubMed=11296296;  
 RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes."  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 DR EMBL: AE006507; AAK33496.1;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 89 AA; 10050 MW; 3C57483BC9F1A10F CRC64;  
  
 Query Match 1.3%; Score 7; DB 16; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 130 VRLLLDD 136  
 DB 81 VRLLLDD 87  
  
 RESULT 30  
 ID 09FRG7 PRELIMINARY; PRT; 94 AA.  
 AC 09FRG7;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Hypothetical 10.2 kDa protein.  
 GN OSJNB0009F04.21.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; Oryza.  
 NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. NIPPONBARE;  
 RC Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,  
 RA Bowman V., Pai G., Bowman C.L., Fujii C.Y., Vanaken S.E.,  
 RA Bismar C.L., Craven B., Utecherback T.R., Khairak H., Feldblyum T.V.,  
 RA Quakenbush J., White O., Salzberg S.L., Fraser C.M.,  
 RT "Oryza sativa chromosome 3 BAC OSJNB0009F04 genomic sequence."  
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC079830; AAG46065.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 94 AA; 10225 MW; 8ADCE0753637E83B CRC64;  
  
 Query Match 1.3%; Score 7; DB 10; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 QY 194 RATILGG 200  
 DB 19 RATILGG 25  
  
 RESULT 31  
 ID 09BRH7 PRELIMINARY; PRT; 109 AA.  
 AC 09BRH7;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE Similar to N-myc downstream regulated.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=LUNG;  
 RA Strausberg R.;  
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC006260; AA06260.1; -  
 DR InterPro: IPR004142; Ndr.  
 SR Pfam: PF03096; Ndr; 1.  
 SO SEQUENCE 109 AA; 11384 MW; BD0096DDBE3E0F CRC64;

Query Match 1.3%; Score 7; DB 4; Length 109;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 416 GSSVSL 422  
 DB 46 GSSVSL 52

## RESULT 32

O9K9X7 PRELIMINARY; PRT; 117 AA.  
 ID O9K9X7  
 AC O9K9X7  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein BH2517.  
 GN BH2517.  
 OS Bacillus halodurans.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=8665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 Fujii F., Hirata C., Nakamura Y., Ogasawara N., Kuhara S.,  
 Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
 halodurans and genomic sequence comparison with Bacillus subtilis.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 DR EMBL: AP001515; BAB06236.1; -  
 KW Hypothetical protein; Complete proteome.  
 SO SEQUENCE 117 AA; 13851 MW; 0F1EB0209230417D CRC64;

Query Match 1.3%; Score 7; DB 16; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 LISLCL 29  
 DB 81 LISLCL 87

## RESULT 33

O9S324 PRELIMINARY; PRT; 123 AA.  
 ID O9S324  
 AC O9S324  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Nitrogenase reductase (EC 1.18.6.1) (Fragment).  
 GN NIFH.  
 OS Paenibacillus polymyxa (Bacillus polymyxa).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Paenibacillaceae; Paenibacillus.  
 OX NCBI\_TaxID=1406;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM230;  
 RX MEDLINE=99354483; PubMed=10425751;

RA Achouak W., Normand P., Heulin T.;  
 RT "Comparative phylogeny of rrs and nifH genes in the Bacillaceae.";  
 RL Int. J. Syst. Bacteriol. 49:961-967(1999).  
 DR EMBL: AJ233995; CAB52206.1; -  
 DR HSRP: P00459; INZC.  
 DR InterPro: IPR000392; NitrogenaseH.  
 DR Pfam: PF00142; fer4\_NIFH; 1.  
 DR PRINTS: PR00091; NITROGNASEH.  
 DR PROSITE: PS00746; NIFH\_FRXC.1; 1.  
 DR PROSITE: PS00692; NIFH\_FRXC.2; 1.  
 KW Oxidoreductase.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SO SEQUENCE 123 AA; 13224 MW; 98B0EEFD7E191F6C CRC64;

Query Match 1.3%; Score 7; DB 2; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 121 YLAERG 127  
 DB 23 YLAERG 29

## RESULT 34

O42163 PRELIMINARY; PRT; 124 AA.  
 ID O42163  
 AC O42163  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE LEA D34-like protein (Mouse-ear cross).  
 OS Arabidopsis thaliana (Mouse-ear cross).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=DRY SEEDS OF A THALIANA ECOTYPE COLUMBIA;  
 RA Raynal M., Grellier F., Laudie M., Meyer Y., Cooke R., Delzeny M.;  
 RL Submitted (Oct-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: Z27037; CA81572.1; -  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SO SEQUENCE 124 AA; 13015 MW; 9A0E1AEF6B3E812 CRC64;

Query Match 1.3%; Score 7; DB 10; Length 124;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 LAERGV 128  
 DB 76 LAERGV 82

## RESULT 35

O9YCW2 PRELIMINARY; PRT; 125 AA.  
 ID O9YCW2  
 AC O9YCW2  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein Ape1150.  
 GN APE1150.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcates;  
 OC Desulfurococcaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 RX MEDLINE=99310339; PubMed=10382966;

RA Kavarabeyasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 RT crenarchaeon, Aeropyrum pernix K1.";  
 RT DNA Res. 6:83-101(1999)  
 DR EMBL: AF000060; BAA80135.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 125 AA; 12756 MW; 6F9EA250896E17E6 CRC64;

Query Match 1.3%; Score 7; DB 17; Length 125;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 259 GLOALCY 265  
 DB 31 GLOALCY 37

RESULT 36  
 Q96MJ2 PRELIMINARY; PRT; 127 AA.  
 AC Q96MJ2;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE CDNA FLJ32293 f19, clone PROST2001739.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 OX NCBI\_TaxID-9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PROSTATE;  
 RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
 RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,  
 RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AK056855; BAB71296.1; -;  
 SQ SEQUENCE 127 AA; 13602 MW; D6E09E1512CD936D CRC64;

Query Match 1.3%; Score 7; DB 4; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 412 KGLTGS 418  
 DB 45 KGLTGS 51

RESULT 37  
 O16513 PRELIMINARY; PRT; 127 AA.  
 AC O16513;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE T05H4.8 protein.  
 GN T05H4.8.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID-6239;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE-94150718; PubMed-7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Cratton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Ritten K., Koopa A., Saunders D., Showkneen R.,  
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Blanchard M.;  
 RT "The sequence of C. elegans cosmid T05H4.";  
 RL Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF016452; AAB66014.1; -;  
 SQ SEQUENCE 127 AA; 15028 MW; D3E9FAF66A3F838 CRC64;

Query Match 1.3%; Score 7; DB 5; Length 127;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 45 TESRHN 51  
 DB 121 TESRHN 127

RESULT 38  
 Q8TVE6 PRELIMINARY; PRT; 130 AA.  
 AC Q8TVE6;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DE Ribosomal protein S6e (S10).  
 GN RPS6A OR MK1446.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyrta; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID-2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE-21927647; PubMed-11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Nalele D.A., Rogozin I.B., Tatuzov R.L., Wolf Y.I., Stetter K.O.,  
 RA Mal'kh A.G., Koonin E.V., Kozayavkin S.A.;  
 RT "The complete genome of hyperthermophilic Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL: AE010437; AAM02659.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 130 AA; 14322 MW; 166CAC9D304F624 CRC64;

Query Match 1.3%; Score 7; DB 17; Length 130;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 128 VRYRL 134  
 DB 111 VRYRL 134

DB 71 VNRVRL 77

RESULT 39

ID 09L101 PRELIMINARY; PRT: 135 AA.

AC 09L101: 09L101

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Similarity to self-Incompatibility protein.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.,

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-COLUMBIA;

RX MEDLINE=20363099; PubMed=10907853;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.

RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,

RT TAC and BAC clones."

RL DNA Res. 7:217-221(2000).

DR EMBL: AP001297; BAB03025.1;

SQ SEQUENCE 135 AA; 15429 MW; 9009221D5AE6804 CRC64;

Query Match 1.3%; Score 7; DB 10; Length 135;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 147 LALDHP 153

DB 37 LALDHP 43

RESULT 40

ID 026067 PRELIMINARY; PRT: 136 AA.

AC 026067: 026067

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Hypothetical protein HP1542.

GN HP1542.

OS Helicobacter pylori (Campylobacter pylori).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI\_TaxID=210;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=26695 / ATCC 700392;

RX MEDLINE=97394467; PubMed=9252185;

RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,

RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,

RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,

RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Wathey L., Wallin E.,

RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,

RA Venter J.C.;

RT "The complete genome sequence of the gastric pathogen Helicobacter

RT pylori."

RL Nature 388:539-547(1997).

DR EMBL: AE000652; AAD08589.1;

DR TIGR: HP1542;

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 136 AA; 14567 MW; ED5652262167A838 CRC64;

Query Match 1.3%; Score 7; DB 16; Length 136;

Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 224 TGSVGE 230

DB 58 TGSVGE 64

Search completed: May 12, 2003, 10:08:48

Job time : 89 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 16:17:26 ; Search time 2856 Seconds  
(without alignments)  
5349.780 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 2713  
Sequence: 1 MRANPKQAMSETISLMKT.....KMKRIAKIISLPITSL 525

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV-xlh  
-MODEL=frame+pn.model -DEV-xlh  
-O/cgn2.1/USPTO/US10066531/runat\_12052003\_091142\_22353/app.query.fasta.1.711  
-DB=GenEmbl -QFMT=fastap -SUFF=irge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human4.0.cdi -LIST=45  
-LOCALIGN=200 -THR.SCORE=Pct -THR.MAX=100 -THR.MIN=0 -ALIGN=40 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10066531.ecgn.1.1.1744\_@runat\_12052003\_091142\_22353 -NCPU=6 -ICPU=3  
-NO\_XLPTX -NO\_MMAPP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: gb\_da:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_in:\*  
5: gb\_ov:\*  
6: gb\_ov:\*  
7: gb\_ph:\*  
8: gb\_ph:\*  
9: gb\_pl:\*  
10: gb\_pl:\*  
11: gb\_ro:\*  
12: gb\_ro:\*  
13: gb\_un:\*  
14: gb\_un:\*  
15: gb\_vl:\*  
16: gb\_vl:\*  
17: em\_da:\*  
18: em\_da:\*  
19: em\_da:\*  
20: em\_da:\*  
21: em\_da:\*  
22: em\_da:\*  
23: em\_da:\*  
24: em\_da:\*  
25: em\_da:\*  
26: em\_da:\*  
27: em\_da:\*  
28: em\_da:\*

29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_in:\*  
32: em\_hcg\_other:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rod:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_vrt:\*  
38: em\_sy:\*  
39: em\_hcg\_hum:\*  
40: em\_hcg\_mus:\*  
41: em\_hcg\_other:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	2713	100.0	329861	1	NMA522491	AL162756 Neisseria
C 2	2683	98.9	10882	1	AE002493	AE002493 Neisseria
C 3	2683	98.9	349980	6	AX044032	AX044032 Sequence
4	1115.5	41.1	10980	1	AE000206	AE000206 Escherich
5	1115.5	41.1	15047	1	D90741	D90741 Escherichia
6	1115.5	41.1	19201	1	D90742	D90742 Escherichia
7	1113.5	41.0	222605	1	AP002555	AP002555 Escherich
8	1106.5	40.8	9255	1	AE006196	AE006196 Pasteurel
9	1106.5	40.8	10190	1	AE005315	AE005315 Escherich
10	1104	40.7	22411	1	AE008749	AE008749 Salmonell
11	1104	40.7	254050	1	AL627269	AL627269 Salmonell
12	870.5	32.1	10979	1	AE011929	AE011929 Xanthomon
C 13	856.5	31.6	10211	1	AE012383	AE012383 Xanthomon
C 14	839.5	30.9	65792	6	AX067456	AX067456 Sequence
15	824.5	30.4	208315	1	AP003017	AP003017 Mesorhizo
16	821	30.3	11330	1	AE008085	AE008085 Agrobacte
17	821	30.3	12593	1	AE009120	AE009120 Arabidac
18	816.5	30.1	347950	1	AE003013	AE003013 Mesorhizo
19	808.5	29.8	12834	1	AE004943	AE004943 Pseudomon
20	777	28.6	1551	6	AX078493	AX078493 Sequence
21	761	28.1	299350	1	SME591786	AL591786 Sinorhizo
C 22	632	23.3	12892	1	AE001456	AE001456 Helicobac
C 23	628	23.1	10631	1	AE000539	AE000539 Helicobac
C 24	492.5	18.2	272345	2	AC090533	AC090533 Mus muscu
25	427	15.7	10626	1	AE003945	AE003945 Xylella f
26	400	14.7	11218	1	AE012113	AE012113 Xanthomon
27	383	14.1	10802	1	AE011644	AE011644 Xanthomon
28	336	12.4	110000	2	LMFCHRI16_00	LMFCHRI16_00
29	332	12.2	110000	2	LMFCHRI18_00	LMFCHRI18_00
C 30	325	12.0	10118	1	BSUB0019	BSUB0019 B. subtilis
31	325	12.0	212610	1	Y29122	Y29122 Bacillus su
C 32	323	11.9	2319	6	AX414863	AX414863 Sequence
33	323	11.9	6273	6	AX416851	AX416851 Sequence
C 34	318	11.7	231450	1	AL596163	AL596163 Listeria
C 35	318	11.7	349980	6	AX413018	AX413018 Sequence
36	318	11.7	349980	6	AX415067	AX415067 Sequence
37	312	11.5	8419	1	AF084042	AF084042 Listeria
38	311	11.5	200050	1	AL591973	AL591973 Listeria
C 39	308.5	11.4	11017	1	AE011624	AE011624 Xanthomon
40	308.5	11.4	307750	1	AP003136	AP003136 Staphyloc
41	308.5	11.4	348650	1	AP003364	AP003364 Staphyloc
C 42	308.5	11.1	270852	2	AC125346	AC125346 Mus muscu
43	296.5	10.9	10394	1	AE006330	AE006330 Lactococc
C 44	296.5	10.9	301450	1	AP003185	AP003185 Clostridi
C 45	296.5	10.9	301450	1	AP003185	AP003185 Clostridi

RESULT 1

## ALIGNMENTS

NMA522491/c  
LOCUS NMA522491 329861 bp DNA linear BCT 04-DEC-2000  
DEFINITION *Neisseria meningitidis* serogroup A strain 22491 complete genome;  
segment 5/7.  
ACCESSION AL162756 AL157959  
VERSION AL162756.2 GI:7380091  
KEYWORDS  
SOURCE *Neisseria meningitidis* 22491.  
ORGANISM *Neisseria meningitidis* 22491.  
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
Neisseria.  
REFERENCE 1 (bases 1 to 329861)  
AUTHORS Parkhill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., Davies, R.M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N., Holtroyd, S., Jagsels, K., Leather, S., Moule, S., Mungall, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Simmonds, M., Skelton, J., Whitehead, S., Spratt, B.G. and Barrall, B.G.  
TITLE Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491  
JOURNAL Nature 404 (6777), 502-506 (2000)  
MEDLINE 20222356  
PUBMED 10761919  
REFERENCE 2 (bases 1 to 329861)  
AUTHORS Parkhill, J.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the *Neisseria* sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk  
COMMENT Notes:  
Details of *N. meningitidis* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/N\\_meningitidis/](http://www.sanger.ac.uk/Projects/N_meningitidis/)).  
Location/Qualifiers  
1..329861  
/organism="Neisseria meningitidis 22491"  
/strain="22491"  
/db\_xref="taxon:122587"  
/feature="serogroup: A"  
194..565  
/gene="NMA1439"  
194..565  
/gene="NMA1439"  
/note="NMA1439", possible lipoprotein, len: 123 aa; unknown, contains a probable N-terminal signal sequence and an appropriately positioned PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
/codon\_start=1  
/transl\_table=11  
/product="putative lipoprotein"  
/protein\_id="CAB84675.1"  
/db\_xref="GI:7380092"  
/db\_xref="SPTREMBL:Q9JU92"  
/translation="MNKTLSTLPAVAILLGGCAAGGNTFGSLDGTGSGSIVKMAVE  
SQCRALNKRESEWRLTALMSAEKQAEWENKICACVAGAEAPNLTGNDVMQMLDPSTR  
NOALALTAFTVSACFKHLVR"  
212..244  
/gene="NMA1439"  
/note="PS00013 Prokaryotic membrane lipoprotein lipid attachment site"  
complement(571..580)  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
592..601  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
misc-feature  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
638..2047  
/gene="thrc"  
638..2047  
/gene="thrc"  
/EC\_number="4.2.99.2"  
/note="NMA1440", thrc, probable threonine synthase, len: 485 aa; similar to many e.g. SW:THRC\_METGL (EMBL:D14071),

thrc, *Methylobacillus glyco*genes threonine synthase (EC 4.2.99.2) (475 aa), fasta scores: E(): 0, 61.9% identity in 475 aa overlap. Contains pfam match to entry PF00291 S.T dehydratase, Pyridoxal-phosphate dependent enzymes"  
/codon\_start=1  
/transl\_table=11  
/product="putative threonine synthase"  
/protein\_id="CAB84676.1"  
/db\_xref="GI:7380093"  
/db\_xref="SPTREMBL:Q9JU91"  
/translation="MKYSTGETAHKPESEVILMGAPDGLMPEHYPOIGREALD  
KWRGLAYPELAEIRLRLVTDTPEDDLILKRTTEAAGFKETTPRTISDGLKIO  
ALNSGPTLAFKMAKQFLGNAPETVLNKGKLNLTGATSGDTSAAETPRATSGDKIO  
VFMSPDGKMSAFQRAQWYSLDQGNINIAVKGMPDDQDITKAVQNDAPAREKHIG  
TVNSINMERIVAOVYFAGYFARKASQNDQGVSPSGNGNCAAGHIAKQMLPVR  
RLIVATNENDVIDEPEFKGAVPRNSATYVSSGSMISKASNEREFVPLMDLDPQ  
ELNTLMAVAGKGNFLPEALDKVKGFGFTGSKTHADRLATITQVVOEOELDPH  
TADGVKVAREVREEREYVCTETALAAFDATIRAVQDVAIAPRAALEGLENLPRV  
QTPNSAIVKGIITQTLA"  
875..1780  
/gene="thrc"  
/note="Pfam match to entry PF00291 S.T dehydratase, Pyridoxal-phosphate dependent enzymes, score 231.80, E-value 1e-65"  
complement(922..931)  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
2094..2888  
/gene="NMA1441"  
2094..2888  
/gene="NMA1441"  
/note="NMA1441", len: 264 aa; unknown"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein NMA1441"  
/protein\_id="CAB84677.1"  
/db\_xref="GI:7380094"  
/db\_xref="SPTREMBL:Q9JU90"  
/translation="WTVPMQETHFSILDELAAQOEAIAIHLADTKVWIRAGRH  
NARWRVALLGWAARYLKLGLVLPVPSLGGEPALIAESRVLRLAAGVAPPELLAVRK  
NALMFGNLEGIPLDPIROEAEAGKADAMGLAIAVHKRQOFLSQAFARNMKSDG  
KNISFLDEDDPEVLTIAOCCARDMLCYIINSTALILKNGSLLEAAAKMGCVLSDOP  
AEIOKLINATVPIPLIRLHPRMGRALPLAASISISLADMP"  
2904..2937  
/note="Stem loop containing DNA uptake sequences: acgat gccgcctgaa gccgcctgaa atata"  
2909..2918  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
complement(2923..2932)  
/note="Core DNA uptake sequence: gccgcctgaa"  
/label=DUS  
3065..3068  
/gene="fpr"  
3076..3852  
/gene="fpr"  
3076..3852  
/gene="fpr"  
/EC\_number="1.18.1.2"  
/note="NMA1442", fpr, probable ferredoxin--NADP reductase, len: 258 aa; similar to many e.g. SW:PFNR\_AZQVI (EMBL:L36319), fpr, Azotobacter vinelandii ferredoxin--NADP reductase (EC 1.18.1.2) (257 aa), fasta scores: E(): 0, 69.8% identity in 255 aa overlap. Similar to NMA1664, fasta scores: E(): 1.3e-30, 34.6% identity in 257 aa overlap. Contains pfam match to entry PF00175 oxidored\_fad, Oxidoreductase FAD/NAD-binding domain"  
/codon\_start=1  
/transl\_table=11  
/product="putative ferredoxin--NADP reductase"  
/protein\_id="CAB84678.1"  
/db\_xref="GI:7380095"  
/db\_xref="SPTREMBL:Q9JRE3"  
/translation="MAAFNRQKVLVSHVHTDAFTFTCTRDSESLRFENGCPVVMGLAV



DGPLMARVSVASANEHELEFFSIKVDGELTSLQHLKVGDDVLISKPTGLVAG  
 DLNPGKHLVLTSTGIGLAPFLSTKDPETLVEQFEKILIVHGRYKRLAYDRFKEL  
 PEHEKLDVLEKELIYPIVSREPEHEHRLDPLWVSKLEFEDIDLPKINPDDBRAML  
 CGSPAMKDRCKVLDVDFGLTVSPKTVGVDILIERAFVDO"  
 3394..3768  
 /gene="tpr"  
 /note="Pfam match to entry PF00175 oxidored.fad,  
 oxidoreductase FAD/NAD-binding domain, score 20.00,  
 E-value 8.4e-05"  
 3866..3875  
 /note="Core DNA uptake sequence: gccgtctgaa"  
 /label=DUS  
 3964..3983  
 /note=">= 90% match to ATTCNNNNNNNGGAAT"  
 /label=DRS3  
 complement(3984..4038)  
 /note="RS103"  
 /label=RS103  
 4039..4058  
 /note=">= 90% match to ATTCNNNNNNNGGAAT"  
 /label=DRS3  
 complement(4038..4038)  
 /label=DRS3  
 4159..4320  
 /gene="NMA1443"  
 /note="NMA1443", len: 53 aa; unknown, lies within a region  
 of unusually low GC content"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein NMA1443"  
 /protein\_id="CAB84679.1"  
 /db\_xref="gi:7380096"  
 /db\_xref="SPTREMBL:Q9J089"  
 /translation="MIYKKNCKNKCYAKENKKRMVFIENGSAIVIRHTGIMKKIHH  
 IDIILPFLA"  
 complement(4509..6980)  
 /gene="NMA1444"  
 complement(4509..6980)  
 /gene="NMA1444"  
 /note="NMA1444, probable P-type cation-transporting  
 ATPase, len: 823 aa; similar to many e.g. SW:COPIA\_ENTHR  
 (EMBL:LI3292), COPA, Enterococcus hirae  
 copper/potassium-transporting ATPase A (EC 3.6.1.36) (727  
 aa), fasta scores: E(): 0, 34.0% identity in 744 aa  
 overlap. Similar to NMA1539, fasta scores: E(): 0, 34.4%  
 identity in 735 aa overlap. Contains hydrophobic, probable  
 membrane-spanning regions. Contains two Pfam matches to  
 entry PF00122 E1-E2 ATPase, E1-E2 ATPases and PS00154  
 E1-E2 ATPases phosphorylation site"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative P-type cation-transporting ATPase"  
 /protein\_id="CAB84680.1"  
 /db\_xref="gi:7380097"  
 /db\_xref="SPTREMBL:Q9J088"  
 Alignment Scores:  
 Pred. No.: 4.82e-200 Length: 329861  
 Score: 2713.00 Matches: 525  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 1 Gaps: 0  
 US-10-066-551-4 (1-525) x NMA522491 (1-329861)  
 QY 1 MetArgAlaAsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThr 20  
 DB 228566 ATGCGCGCCACCCCAACACAGCAATGCCGTGAAACCAATCCCGATGAGAAACA 228507  
 QY 21 ArgSerLeuIleSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
 DB 228506 CGACGCTAATATTCCTTTATGCTCTCTCTGTTCAATGTTCTTCATGTTGCCCCCA 228447

QY 41 LeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsn 60  
 DB 228446 CTGGAGAACGAGCAGGAACCCCTCATTTCAATACTTCCAAACCCGCGCGGAGAAC 228387  
 QY 61 IleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsn 80  
 DB 228386 ATCCCTCAATCCGCGACACCCCTCATACCAACGGGCTATCCGATATCTGTGTGAAC 228327  
 QY 81 AspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeu 100  
 DB 228326 GACCCCGAGAACCTTTGGCGCGCGCGCGCTTATCGAATCTCCGGAACACAGCTTC 228267  
 QY 101 AspLeuGlnTyrTyrIleTyrPArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuVal 120  
 DB 228266 GATTTCGCAATCACTACATCTGGCCCAACGACATTTCCGCGGACTGCTGTTCAACCTCTG 228207  
 QY 121 TyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArg 140  
 DB 228206 TACCTTGCGCGAGACGCGGTGTGCGGTACGCTGTGTGAGCAGCAACACAGCGCG 228147  
 QY 141 GlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPhe 160  
 DB 228146 GGATTTGAGCAGCTCTGCTGCGCTTCGACAGCATCCCAATATCGAATGCGCTGTTTC 228087  
 QY 161 AspProPheValLeuArgGlySerTrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeu 180  
 DB 228086 AACCCCTGCTCTTACGAAATATGGCGCGCTGGCTACTGACCGCACTTCCCGCGCTC 228027  
 QY 181 AsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGly 200  
 DB 228026 AACCGCGCATGCAACAATCTTATCCGCGCAGCAACCGCGCCACCATATCTGCGGGA 227967  
 QY 201 ArgAsnIleGlyAspGluTyrPheLysValGlyGluAspThrValPheAlaAspLeuAsp 220  
 DB 227966 CGCAATATCGCGCAGCAATATCTTCAAGTGGAGAGACCCCTTTCGCGCAGCTGGAC 227907  
 QY 221 IleLeuAlaThrGlySerValValGlyGluValSerHisAspPheAspArgTyrTrpAla 240  
 DB 227906 ATCTCGCGCCCGCGCAGCGCTGCGCGGAAGTATCGACACATTCGACCCCTACTGGGCA 227847  
 QY 241 SerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGlyLeu 260  
 DB 227846 AGCCATTTCCGCCCAACAGCGCAGCGATCATCGCGGCAACATCGGCAAGGCTT 227787  
 QY 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThr 280  
 DB 227786 CAAGCACTCGGATACAAACGAGCAACGTCAGACACCGCTCTGCGCTACCGGAAACC 227727  
 QY 281 ValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerVal 300  
 DB 227726 GTGCAACAGTCCGCCCTCTACCAAAAATACAGACAGACGATCGATGCGAGAGCTC 227667  
 QY 301 GlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysPro 320  
 DB 227666 CAACACCGCTCATACAGCGAGCAGCCCTGCAAAAGACTCCAGCCGCGCAACACCG 227607  
 QY 321 ProIleAlaGlyArgLeuGlnAspAlaLeuLysGluProGluLysSerValTyrLeuVal 340  
 DB 227606 CGCATTTGCCGGGGGCTGCAAGACGCGCTCAACAGCGCCAAACACGCTATCTGTGTT 227547  
 QY 341 SerProTyrPheValProThrLysSerGlyTyrAspAlaLeuAlaLysLeuValGlnAsp 360  
 DB 227546 TCACCTATTTGTCGCCACAAATATCCGACAGACGACACTGGCAAAATCGTGACAGAC 227487  
 QY 361 GlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaValHis 380  
 DB 227486 GGCATGACCTTACCGCTTACCAACCTCTACAGCGCAGCAGCTTCCGCGCTTCAT 227427  
 QY 381 SerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeu 400  
 DB 227426 TCCGGCTATGTCAAATACGAAACACGCTCTCAAAACCGGATCAACTCTACGACCTG 227367

OY 401 GINPPOASHIALAVALPRALATHRLYASPLYSGLYLeuthrGlySerValThr 420  
 DB 227366 CAACCAACACCATGCGCTCCCTCCACAAAAGACAAAGGCGTCCGACGCTCCGTAACC 227307  
 OY 421 SerLeuHISIALYsthrPheIleValaspGLYstargILlePheIleGlySerPheasn 440  
 DB 227306 ACCCTGCATGCCAAACCTTCAATTGTGGACGCAAGCATCTTCATCGCTCATTCAC 227247  
 OY 441 LeuaspProArgSerAlaArgLeuasnThrGluMetGlyValValIleGluSerProLys 460  
 DB 227246 CTCGACCCCGCTCCGACGCTCATACTGAATGGCGGTGTATTCGAAGCCCA 227187  
 OY 461 IlealaglGluMetGluThrThrLeuAlaaspThrSerProGluThrAlaTyrargVal 480  
 DB 227186 ATCGGAGACAGATGAGGCGACCCCTGCGATCTCACCCGAAATACGCTACCGCGTT 227127  
 OY 481 ThrLeuaspArgHISasnArgLeuGlnThrPHisaspProAlaThrArgTyrThrTyrPro 500  
 DB 227126 ACCCTGCAGACGACACCGCTCGCATGCGACAGATCCCGCCGCAAAACCTACCGC 227067  
 OY 501 AsngluProGluAlaLysLeuThrPlystargILleAlaLysIleLeuSerLeuLeuPro 520  
 DB 227066 AACGAAACCGAAGCAAACTTGGAAACGATCGCGCAAAATCCTATCCTGCTGCC 227007  
 OY 521 IleGluSerLeuLeu 525  
 DB 227006 ATGAAAGTTATTA 226992  
 RESULT 2  
 AE002493/c 10882 bp DNA linear BCT 25-MAY-2000  
 LOCUS Neisseria meningitidis serogroup B strain MC58 section 135 of 206  
 DEFINITION Of the complete genome.  
 ACCESSION AE002493 AE002098  
 VERSION AE002493.1 GI:7226674  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 . Neisseria meningitidis MC58.  
 Neisseria meningitidis MC58  
 Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae;  
 Neisseria.  
 1 (bases 1 to 10882)  
 Tettelein,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,  
 Clifton,H., Clark,E.B., Cotton,M.D., Uterback,T.R., Khouri,H.,  
 Qiu,H., Yamachyan,J., Gill,J., Scarlato,V., Maignani,V.,  
 Pizzo,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
 Rappunli,R. and Venter,J.C.  
 Complete genome sequence of Neisseria meningitidis serogroup B  
 strain MC58  
 JOURNAL Science 287 (5459), 1809-1815 (2000)  
 MEDLINE 20175755  
 PUBMED 10710307  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 10882)  
 Tettelein,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,  
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,  
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,  
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,  
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,  
 Clifton,H., Clark,E.B., Cotton,M.D., Uterback,T.R., Khouri,H.,  
 Qiu,H., Yamachyan,J., Gill,J., Scarlato,V., Maignani,V.,  
 Pizzo,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,  
 Rappunli,R. and Venter,J.C.  
 Direct Submission  
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA  
 TITLE  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1. 10882  
 /organism="Neisseria meningitidis MC58"  
 /strain="MC58"  
 /db\_xref="taxon:122586"

gene  
 CDS  
 /note="serogroup: B"  
 complement(135..1661)  
 /gene="NM01434"  
 complement(135..1661)  
 /gene="NM01434"  
 /note="similar to GB:000096 SP:p75919 PID:1787284 percent  
 identity: 68.05; identified by sequence similarity;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="cardiolipin synthetase family protein"  
 /protein\_id="AA041795.1"  
 /db\_xref="GI:7226675"  
 /translation="MKTRLSLSLCLLCSSSWLPPEERTESRHNFTSKPRLDNT  
 LQIRTPHTNGISDIYLLNDPHEFAAARALIESAHSLDIQYIWRNDISGRLEFLN  
 VYLAERGVRLVLLDDNNTRGLDDLLALDSHPNIEVRLFPYLRKRALGYLTFD  
 PRLNRHMKSTFADNRATITIGRNIGDEYKVGSDYVAFADILATGSGVGSNDP  
 DRYASASHANATRIIRSGDIDGKIGLQALGYNDERSRALIKREYVESPIYKIQTC  
 CIDWQSVTRLSDDPARGLDNRKRPPIAGRLDALOKPEKSVILVSPYVPTKSGT  
 DALAKLVODGIDVYVLTNSLDATDVAAAVHSYVKRPLKAGIKLYELQPNHAPAT  
 KDKGLTSSSVTSLHAKFTIVDGKRIETFSFNLDPKRSARLNTGEMGVIESPIAEQMER  
 TLADTTPAYAVRVTLDRHNRRLQWHPDPRKTYPNPEPAKLWKRIAKILSLPLEGL  
 ."  
 complement(1773..3179)  
 /gene="NM01435"  
 complement(1773..3179)  
 /gene="NM01435"  
 /note="similar to GB:L10328 SP:p31474 PID:290604 GB:000096  
 PID:1790195 percent identity: 66.00; identified by  
 sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="drug resistance translocase family protein"  
 /protein\_id="AA041796.1"  
 /db\_xref="GI:7226676"  
 /translation="MEKNTISARAPSPWLPILLIARPMQMDATILNTAIPETADL  
 NESPLDMOLAVISTYITVALLIPISGVIADRGTRKYVFGSIAFPMDSALCAAGSL  
 FELTISRVOGIGSMVLPITRLTLRYDSSKLNAINIVAMPILISPAACGLILGAGY  
 LVEYASWMIPLNLPIGLIGTILGRNIMPDIKSNISLDKRGVLIIESAACGLILSA  
 ELSHALPPEYFALLPLCGGLFARRYFERHMKYASKPIYSADLFLIRPRLGAGLFS  
 RLGISSEIPMLMFOIAFGFASLGGWLVAPALSSLVKPLFLAPLRKRGYRVLL  
 WNTKLAAVIMLIPDGNISPLIMVFLSLAIGANSIOPFSAMNTITLADLRPOOTS  
 GNSIMLAAMQOLAISGVIAGAILIKNMFILPILPASSGLHSARMTLLISGITLASSLV  
 EKRLHVSQGTNLTRNTPS"  
 364..4243  
 /gene="NM01436"  
 364..4243  
 /gene="NM01436"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA041797.1"  
 /db\_xref="GI:7226677"  
 /translation="MSATIPPKIIRYDSNPPTYVFFGTGVLDLPEMPAGMDATILIEO  
 QGIRVHPMAQSCGDPAYSSGHPTEADVAKADLDFPENMPTIVPSGSGGMMKHH  
 WPTLFKSEYERADVACGRRIETHTFLAIGFPEDEKGEPLKVAVHTSCAAREMY  
 HLSMDGLDGMENVERIVHDHESCCGCGGFFSKODISAMNTDYAAALKERGATE  
 IISDCCGMNIIGKIKADEPDMRPKHIAFLIERTGKA"  
 4240..4941  
 /gene="NM01437"  
 4240..4941  
 /gene="NM01437"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA041798.1"  
 /db\_xref="GI:7226678"  
 /translation="MSARENILAKLAKADALPMEPAVEDYIRREMGYSWGSVEVRLKH  
 WAAAMRAVTEIETIYVTKSNMNVQFERAAEGGLKLLIPLATEHGOIARALADSNI

PIAFERIDTWKTEFTFNIDAGSGAOCGIARTGLMFFSSPEERLTSLVPVHCL  
EDTSKMYNEFNHNAVEGEKIVENGMPITVNLISGSKTADQLTLAYAGHGRDLVILA  
ILPDHISPADLEENA"  
4938..6392  
/gene="NM01438"  
/note="conserved hypothetical protein; identified by  
Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAFA1799.1"  
/db\_xref="GI:726679"  
/translation="MTTOTIKFHKMPFKNONAIISLODKPLRSKTRAMDMTKRK  
AVLTDEELOSRLDCEHVRKRSLSKIPALLEOLEENITKIGVYVHNAETTECCOII  
HDITAKNGKLMVKGSMVSEIELNLLEKGIKAVESDGEFIVQAGKPHIVIA  
PAIHRTKEOVSELFHONIGPLTDVDQLGFAKALADIVSTADVGISGVAFVAET  
GTLCLVNEGNGRLSTVPPVHIAITGIEKVAKSIDIPLYSLPSAIGONITVE  
NMTGPRSEELDGPOMHLVLDNGRQAVAEQOMRTLOCICGACMNHCPYTRI  
GGAAYGTTPGPIGEIISPHLIGDTRDLPTACTMCGACVGCVPRIPIEOMORLR  
VEORSPTEYVPHIRGOGASHTRGEOAMARTFGISGSKTYAFGMAATKRFNLLP  
RKOLGTONRPMKPAKTIHLEMAEKROEQA"  
/complement(6575..7060)  
/gene="NM01439"  
/complement(6575..7060)  
/note="similar to GB:J02732 SP:P12044 PID:143364  
PID:2239295 GB:AL009126 percent identity: 76.29;  
identified by sequence similarity; putative"  
/transl\_table=1  
/codon\_start=1  
/product="phosphoribosylaminoimidazole carboxylase,  
catalytic subunit"  
/protein\_id="AAFA1800.1"  
/db\_xref="GI:726680"  
/translation="MIQIGIMSGNSDMPVNRQAQPLEEGVEYEARVVAHRTPLD  
MFOYAEVFAARGAIIAGAGAAHLPQMVAKTIVLGVPVSKYLRGDSLSIY  
OMPKGVPAFPAIGEGAAANAALFAVSMLANETDELAKLADFPKADQAVLNMELQ  
I"  
/complement(7108..7821)  
/gene="NM01440"  
/complement(7108..7821)  
/gene="NM01440"  
/note="hypothetical protein; identified by Glimmer2;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAFA1801.1"  
/db\_xref="GI:726681"  
/translation="MKIKLPLFIILMSVASCASVPVPAQSOTEMSTRNADGIPY  
PVPTLDRDLDEKRIYRLSNEVETLNGKVALBHAKTHSSGRAVYOKLDRKKEH  
LNTGGSASATHTVETAOINQALKHRSKGFSAASLTKADGGDSIQRSMYLL  
LOSRAVGNCEVIEIGRYANRKPDPJAEAMFKIGECQYRLQKDIARATWSLI  
QTYGSPAAKRAAAVRRK"  
/complement(7821..8489)  
/gene="NM01441"  
/complement(7821..8489)  
/note="similar to PID:1001734 PID:1001730 percent  
identity: 59.73; identified by sequence similarity;  
putative"  
/codon\_start=1  
/transl\_table=1  
/product="0-methyltransferase, putative"  
/protein\_id="AAFA1802.1"  
/db\_xref="GI:726682"  
/translation="MTTHLSNVAPDLQNYLNAIGEHPVLTRLREKTHGRMGKMAI  
ARQOAVLWMLAKLIRAEKYLEIGVTGYSSTLALALPEHGRITACDINTFTDAR  
QVWNEACVAHKISLHLOPALTIDDLIAGSAGSYDALIDADKPPPTQYTERCKLY  
RGGIILIDNILLNGRVAREASADPSVGLKDFNONLPDPRIPTLTPVSGDLTL  
LKR"

gene 8829..10805  
/gene="NM01442"  
8829..10805  
CDS  
/note="similar to GB:M29686 SP:P14160 GB:S56948 PID:153657  
percent identity: 54.18; identified by sequence  
similarity; putative"  
/codon\_start=1  
/transl\_table=1  
/product="mismatch repair protein MutL"  
/protein\_id="AAFA1803.1"

Alignment Scores:

Pred. No.:	1,296-199	Length:	10882
Score:	2683.00	Matches:	519
Percent Similarity:	59.43%	Conservative:	3
Best Local Similarity:	98.86%	Mismatches:	3
Query Match:	98.89%	Indels:	0
DB:	1	Gaps:	0

US-10-066-551-4 (1-525) x AE002493 (1-10882)

QY 1 MetArgAlaAsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThr 20  
1712 ATGCCGCGCAACCCCAACACAGGCAATGCCGTGAAACATATCCCTGATGAACA 1653

QY 21 ArgSerLeuIleSerLeuLeuCysSerCysSerSerTrpLeuProPro 40  
1653 CGCAGCTTAATTCCTTTATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1593

QY 41 LeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsn 60  
1592 CTGGAAGAAGCGAGGAAAGCCGTCATTTCATCTTCATTCGAAACCGCTCCGCTGGAACA 1533

QY 61 IleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleThrLeuLeuAsn 80  
1532 ATCCCTGCAATTCGGCGACACCCCTCATCCAAAGGGGCTATCCGATATCTATCTGTGAAC 1473

QY 81 AspProHisGluAlaPheAlaIleArgAlaIleuIleGluSerAlaGluHisSerLeu 100  
1472 GACCCCGCAGAGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413

QY 101 AspLeuGlnIleThrIleArgHisThrProHisThrAsnGlyLeuSerAspIleThrLeuLeuAsn 120  
1412 GATTTCGCAATTCATCTATCTGCGGCAACGACATTCCTCCGAGGCTCTGTCAACTCGTG 1353

QY 121 TyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAspAspAsnThrArg 140  
1352 TACCTTCCCGCAGAACCGCGGTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1293

QY 141 GlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPhe 160  
1292 GGATTGGACGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1233

QY 161 AsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeu 180  
1232 AACCCCTTCGCTTACGAAATATGGCGGACCTCGGTACTGACCGACTTCCCGCGCTTC 1173

QY 181 AsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGly 200  
1172 AACCCCGCATGACACAAATCTTTACCCCGCAGAACCGCGCACCCTACTCTGCGCGGA 1113

QY 201 ArgAsnIleGlyAspGluTyrPheLysValIleGlyLysPheValPheAlaLeuAsp 220  
1112 CGCAATATCGGCGAGATCTTCAAGTGGGTGAGACACCGTTTCCGCAATTTGGAC 1053

QY 221 IleLeuAlaThrGlySerValValIleGlyLysValSerHisAspPheAspArgTyrTrpAla 240  
1052 ATCCCTGCGACCGGAGCGCTGCGGCGAAGATGACAGACTTGACCGCTACGCGGA 993

QY 241 SerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeu 260  
992 ACCCATTCGCGCCACAGCCACGCGGACATCCGACGCGGACATCGGCAAGGCTCTT 933

QY 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThr 280  
 DB 932 CAAGCACTCGATACAGACGAAACGTCACACGCGCTCCGCTACCGGCGAAAC 873  
 QY 281 ValGluGlnSerProThrTyrGlnIleGlnThrArgIleAspTrpGlnSerVal 300  
 DB 872 GTCCGACAGTCCGCTCCCTACCAAAATACAGACAGATGCTGACTGGAGAGCGTC 813  
 QY 301 GlnThrArgLeuIleSerAspAspProAlaTyrGlyLeuAspArgAspArgTyrPro 320  
 DB 812 CGAACCCGCTCATCAGGAGACACCTCGAAAGAGACTGACCGGACCGCGCAACCG 753  
 QY 321 ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuVal 340  
 DB 752 CCGATTGGCGGGGCGCTCAACAGCCGCTCAAAACAGCCGAAAGAGCTATCTGCTT 693  
 QY 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAsp 360  
 DB 692 TCACCTATTTCGTTCCACAAATCCGACAGACGACGCTGGCAAACTGGTGCAGGAC 633  
 QY 361 GlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaHis 380  
 DB 632 GGCATAGACGTTACGCTTCTGACCAACTCGTCGACGCGACCGACGCTTCCGCTCAT 573  
 QY 381 SerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeu 400  
 DB 572 TCCGCTATGTCAATACGAAACCGCTGCTCAAGCGCGCATCAACTACAGAGTG 513  
 QY 401 GlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThr 420  
 DB 512 CAACCAACCAATGCGCTCCGCGCACAAAGACAAAGCGCTGACGCGACGCTCCCTAAC 453  
 QY 421 SerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsn 440  
 DB 452 AGCGTGCACGCGCAAACTCTATGTGTGACGCGCAACGCACTTCTATGCTGCTTCAAC 393  
 QY 441 LeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerProLys 460  
 DB 392 CTCGACCCCGCTCCGCGCTCAACACCGAAATGGCGCTGTATGAAAGCCCGCAA 333  
 QY 461 IleAlaGlnIleMetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgVal 480  
 DB 332 ATCCGACAGACATGAGACGACCTTCCGCTACCAACCGCGCTACGCTACCGCGCT 273  
 QY 481 ThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLysThrTyrPro 500  
 DB 272 ACCCTGCACAGGACAAACCGCTGCAATGGACATCCCGCACCGCAAAACCTAACCG 213  
 QY 501 AsnGluProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSerLeuLeuPro 520  
 DB 212 AACGAAACCGGAAAGCTTTGAAAGCATCCGCGCAAAATCTATCCCTGCTGCC 153  
 QY 521 IleGluSerLeuLeu 525  
 DB 152 ATAGAAAGCTTATTAT 138  
 RESULT 3  
 AX044032/c 349980 bp DNA Linear PAT 24-NOV-2000  
 LOCUS AX044032 Sequence 111 from Patent WO0066791.  
 DEFINITION AX044032  
 ACCESSION AX044032  
 VERSION AX044032.1 GI:11342916  
 KEYWORDS  
 SOURCE Neisseria meningitidis.  
 ORGANISM Neisseria meningitidis  
 Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
 Neisseria.  
 REFERENCE 1 (bases 1 to 349980)  
 AUTHORS Piazza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C., Masiarant, V., Galeotti, C., Mora, M., Ratti, G., Scariselli, M., Scariselli, V., Rappunli, R., Frazer, C. M. and Grandi, G.  
 TITLE Neisseria genomic sequences and methods of their use

JOURNAL Patent: WO 0066791-A 111 09-NOV-2000;  
 CHIRON CORPORATION (US) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)  
 FEATURES Location/Qualifiers  
 source 1. 349980  
 /organism="Neisseria meningitidis"  
 /db\_xref="taxon:487"  
 /note="sequence too long, cut in 8 pieces.-seq 1: 1 to 349980 349980 bases-seq 108: 300001 to 649980 349980 bases-seq 109: 600001 to 949980 349980 bases-seq 110: 900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980 349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq 113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to 227325 172325 bases"  
 BASE COUNT 86771 a 92803 c 86340 g 84066 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.16e-197 Length: 349980  
 Score: 2683.00 Matches: 519  
 Percent Similarity: 99.43% Conservative: 3  
 Best Local Similarity: 98.86% Mismatches: 3  
 Query Match: 98.89% Indels: 0  
 DB: 6 Gaps: 0  
 US-10-066-551-4 (1-525) x AX044032 (1-349980)  
 QY 1 MetArgAlaAsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThr 20  
 DB 275022 ATCGCGCGCAACCCCAAAACAGCAGCATGCCGTGTGAACCATATCCCTGATGAAACA 274963  
 QY 21 ArgSerLeuIleSerLeuLeuCysLeuLeuLeuCysSerCysSerSerTyrLeuProPro 40  
 DB 274962 CGCAGCTTAATTTCCCTTTATGCTCTCTCTGTTATGTTCTTCAATGCTTCCCGCA 274903  
 QY 41 LeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsn 60  
 DB 274902 CTGACAGACGAGCAGGAAGCGGTGATTTCAATCTTCAAAACCGTCCGCTGGACAC 274843  
 QY 61 IleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsn 80  
 DB 274842 ATCCGCAAAATCCGCGACACCCCTCATATACCAAGGCGTATCGAATATCTGTTGAC 274783  
 QY 81 AspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeu 100  
 DB 274782 GACCCCGACGAAGCTTTGCGCGCGCGCGCCCTTATGCAATCTGCGGACACAGCCTC 274723  
 QY 101 AspleuGlnTyrTyrIleTyrPheArgAsnAspIleSerGlyArgLeuPheAsnLeuVal 120  
 DB 274722 GATTGCAATACATCTGCGCGACGACATTTCCGCGAGGCTGCTTCAACCTGCTG 274663  
 QY 121 TyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuLeuAspAspAsnThrArg 140  
 DB 274662 TACCTTGCGCGAAGACGCGGTGGCGCTACGCGCTGTGGACGACAAACACAGCGGC 274603  
 QY 141 GlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPhe 160  
 DB 274602 GGATTGACAGACCTCTCTGCTCCCTCGACAGCATCCCAATATGCAAGTCCGCTGTC 274543  
 QY 161 AsnProPheValIleuArgLysTyrPheArgAlaLeuGlyTyrLeuThrAspPheProArgLeu 180  
 DB 274542 AACCCCTTGCTTACGAAATGCGCGACCTCGCTACCTGACGAGCTTCCCGCGCTC 274483  
 QY 181 AsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGly 200  
 DB 274482 AACGCGCGCATGCAACAAATCTTATACCGCGCAACACCGCGCACCATCTCGGGGA 274423  
 QY 201 ArgAsnIleGlyAspGluTyrPheLysValGlyLysPheThrValPheAlaAspLeuAsp 220  
 DB 274422 CGCATATCCGCGACGAGATCTTCAAGTCGTAGGACACCGCTTTCGCGCATTTGGAC 274363  
 QY 221 IleLeuAlaThrGlySerValValGlyValSerHisAspPheAspArgTyrTyrAla 240  
 DB 274362 ATCTCGCGACCGCGACGCTGCTGCGGAGATATCGACAGCACTTGCACCGCTACTGGGCA 274303



```

AMWPKEMKSGSVELEIPTNDETNDNI VAAWTPDOLPEGKEMKRYTTTSSRBDP
KLAHPDANAAQQRIRNSYBGDKVMSLIRQDPTIAAVDTGEMKKLPEDTPVTAOTS
IGDGELEIVESTVRYPTKGMRLNRYKRWKDAKKTTEKTAALVGNADDTLSETSYOLP
ANE".
5121..7664
/gene="mdoh"
/note="b1049"
5121..7664
/gene="mdoh"
/function="enzymic; Osmotic adaptation"
/ncore="0847; 99 pct identical to MDH_ECOLI SW: P313137"
/codon_start=1
/transl_table=1
/product="membrane glycosyltransferase: synthesis of
membrane-derived oligosaccharide (MDO)"
/db_xref="GI:1178787"
/translation="MKNTTEYIDAMPAAESEKAAALPKTDIRAVHQAOLDAEHRTWARED
DSQPSKATLEQAMPDPSLADGCLIDDEGRDLAMPPEARSSMFPDPMPTNVGFRF
MDRAGMDVPRYLAIRLTKEGSEBOKMTVTGIRYLLITLTAQVATVAWVKITL
PYRGMALINMDWVGODIWSPEMLPYMLOGLTILPAVLFCWASVAGFPMALMGLO
LILRDKKSYTSASTVDEDELPENHRAILMPICNDUNVRVAGLRATWESKATGANK
HEDYITLSDSTNPDICEAEKAMELEIAVEGEGDITRRKRYKRSKSGNIDFCRR
WGSQYITWVLDHDSVWTGDCGLGVLRAELMPPNAGIIQSSPKASGMDTLYARQCPA
TRVYGRPLTFLWQLGESHVGNAAIIRVPLTEHRCALPLGEESFAGSILSHDF
VEALNRAACMGWIAIYDLPJSGEELPRLIDELKRDRMGNGMLNMFRLFLVGMHP
VHRVFLTGWYSLTSLAPLMEFLESTALAOVNAHLETPQYLQPROLFPWPQWRPEL
ALAEFASITVLELPLKILSTILWCKGTKEGFMRYLSTLLEVLFSVLLAPVRLMF
HRTVVSAPFLGFWVNSVNSPRDDTSNWEAKRRGSGOLILGIVAGVAMGMDLRFLF
MLATIVSILISFPVYSISKATVGLGRTKMKPLIIPPEYSIPVYLVDTDRLEMDRO
KSLDDGMHAEVNPSPNALATALATRRHSAVLEIADRIVAGOLENTEPEKLRNDRR
LLVLSDPVTAARLHFRWNSPERSYSWYEGILNPLNLRKPDAAQ"
6762..7772
/ncore="REP (repetitive extragenic palindromic) element;
contains 2 REP sequences"
7689..7717
/ncore="factor Sigma70; predicted +1 start at 1112689"
7787..7816
/ncore="factor Sigma70; predicted +1 start at 1112788"
7837..8064
/gene="ycek"
/ncore="b1050"
7837..8064
/gene="ycek"
5.31e-77 Length: 10980
1115.50 Matches: 237
60.44% Conservative: 90
43.81% Mismatches: 159
41.12% Indels: 56
1 Gaps: 9
(25) x AE000206 (1-10980)
-----Trrleprprrprrleuglnlgrhthgluser--- 47
||||||| : : : : : : : : : : : : : : : : : :
AGTACTGGGGTTTATGGTATCCCTC-GTGGGAGGGGGGTGAATCGGAG 489
-----Argleuaspanslle 61
||||||| : : : : : : : : : : : : : : : : : :
TTCAGAAATTTATACCGGTACAGCTTTACCCGACAGGATACTTGTGCT 549
-----Argleuaspanslle 61
||||||| : : : : : : : : : : : : : : : : : :
Tghlsthprhsthlnasn----- 71
||||||| : : : : : : : : : : : : : : : : : :
GGAAAGACGCCACCTCTACGAAAGACTCCTTACCCACAGAGAGATGAT 609
-----Glyleuser 74
||||||| : : : : : : : : : : : : : : : : : :
GGCTGGGAGAGCGCGGTCTCCACCTGTGTTCCGACACATCCCGGTACGT 669
-----Glyleuser 74
||||||| : : : : : : : : : : : : : : : : : :
Tleuasnasprrhsthlnalaphaalalalargalalaleuilegln 94
||||||| : : : : : : : : : : : : : : : : : :

```

QY	447	Argleu	snrtrpglumergl	valval1	legius	serprolys	ilial	glugln	metglu	466
										466
Db	1789	TTGCTCA	TACTGA	AAATGG	CTTCTG	TATG	AGAC	CCGAAC	CGCTG	1848
QY	467	ArgThr	leuAla	aspPrh	Ser	proGlu	tyrAla	larTr	argVal	1848
										486
Db	1849	AAACGCT	TATTCAC	AGAC	CGATGAT	CGCGCT	CGGAC	GTCCG	CTGAC	1908
QY	487	Arg	leuGln	trpHis	aspPro	Ala	thr	arg	lys	504
										504
Db	1909	CGGATCA	ACCTGG	GGTTG	ATGCAT	GCAC	AAAG	AAGAT	TATTC	1968
QY	505	Ala	lysLeu	trpPly	argAla	Ala	Ala	lysIle	serLeu	524
										524
Db	1969	ACCACTT	TCTGGA	AGCGGG	GTATG	GTCA	GACTG	CGCTG	CATAT	2028
QY	525	Leu	525							
Db	2029	TTG	2031							
RESULT 5										
LOCUS	D90741									
DEFINITION	D90741	Escherichia coli genomic DNA. (23.7 - 24.0 min).								
ACCESSION	D90741	AB001340								
VERSION	D90741.1	GI:1651509								
KEYWORDS		Complete and shotgun sequencing; csGC; csGF; csGE; csGD; csGB; csGA; ycdB; cIs; nov; mdog.								
SOURCE		Escherichia coli								
ORGANISM		Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.								
REFERENCE		1 (sites)								
AUTHORS		Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., Yano, M. and Horiiuchi, T.								
TITLE		A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map								
JOURNAL		DNA Res. 3 (3), 137-155 (1996)								
MEDLINE		97061202								
REFERENCE		2 (sites)								
AUTHORS		Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.								
TITLE		The systematic sequencing of the Escherichia coli genome in Japan								
JOURNAL		Unpublished								
REFERENCE		3 (bases 1 to 15047)								
AUTHORS		Mori, H.								
TITLE		Direct Submission								
JOURNAL		Submitted (29-JUL-1996) Hirotsu Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5, Takayama, Ikoma, Nara 630-01, Japan								
REFERENCE		E-mail: imori@gtc.ais-t-nara.ac.jp, Tel: 81-7437-2-5660, Fax: 81-7437-2-5669								
COMMENT		Collaboration Information:								
		Project:								
		The Japan E. coli genome DNA sequencing project								
		Group:								
		The Japan E. coli genome DNA sequencing group								
		Members: (1995.4 - 1996.3)								
		Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,								

Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,  
Sampel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,  
Yamamoto, Y., and Yano, M.

# Headed by:

Name: Takashi Horinchi  
Address: National Institute of Basic Biology, Okazaki, 444, Japan  
E-mail: kishiorienb.d.ac.jp  
Information operator:

Name: Hirotada Mori  
Address: NARA Institute of Science and Technology,  
Ikoma, 630-01, Japan

E-mail: hmori@cgic.aist-nara.ac.jp

# URL:

The Japan E. coli genome database  
http://sw3.aist-nara.ac.jp.  
Location/Qualifiers

## FEATURES

### source

1. 15047  
/organism="Escherichia coli"  
/strain="K12"  
/db\_xref="taxon:562"  
/map="23.7-24.0 mln"  
/clone="Kohara clone #231"  
/note="Nucleotide position 1097101-1112147 from the  
initiation site of *thra* (0 min.). This clone is from  
Kohara lambda miniset library."  
1. 276  
/note="ORF\_ID:0230#5"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAA35822.1"  
/db\_xref="GI:4062605"  
/translation="DLPRSENALSTFPKPLMLRLFLIVIEWMTLVSLATLVE  
WRWPEAFELLCYPCILTYFFYCFRLHNDPMACDMYFRMGHNK"  
1152. 2090  
/note="ORF\_ID:0231#1  
similar to SwissProt Accession Number P45637"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical 33.0 kd protein in prob-proA  
intergenic region."  
/protein\_id="BAA35823.1"  
/db\_xref="GI:4062606"  
/translation="MDIFYHPTFDTQWIMIALRKRAIPQARVRAKSGDNDADYALV  
WHPVEMLAGRDILKAFALGAGVDSILSKQAHPEMLNPSVPLFRLDETCGDEMOEY  
AVSGVILHMRFRPDYRQONSSHNPRLPEYHREDPTGILGAGVGSVNASLDTWR  
PLRCWSTRKRMKPGVQSGRRELSAFUSQCKVLINLNPETVGIINQDLKRLPD  
GAYLNLNARGVHVEDDLALDSGKVGAMLDVFNRPPLPESPPLMWHPPRVITTPHY  
AATRPAAVEYISRTIAQLKEGRCVGVDRARGY"  
2145. 2882  
/note="ORF\_ID:0231#2"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAA35824.1"  
/db\_xref="GI:4062607"  
/translation="MYVDLHMTVAHVASTSDYIAQAKOKIKLFAITDGPDM  
EDAPHHMFIMRIMPRVYDVGILRGIEANIKVNDGIDCGKMPDSDLIAGFHE  
PVFAPHDKATMTQAMIAITASGVNHIISHPCNPYEDIVKAVAAKHOVALINNS  
SFLSKRGSSENCREVAAYADAGGVALGSDSHTAFTMGFECECLKLDVDPPEP  
ILNVSPRLNLFLESRGAPTAIEFADL"  
2906. 3460  
/note="ORF\_ID:0231#3  
similar to PIR Accession Number F64035"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical protein H11543"  
/protein\_id="BAA35825.1"  
/db\_xref="GI:4062608"  
/translation="MNEPSILRVLSIGSYVROPDPLVPLFTLIRGKLAANPLPQ  
DELLRLQSCDMTOVSADYALFIDGECANPVPYSAMVEGATBAEVRAPLSEGMEL  
ADTPADHIGTILLAAINLEDOSTDESEBALTLESEYLLPWCAGFLGVEAHATTPEM  
RTMAPLRDAISAMDELEEDSEE"  
3514. 4053  
/note="ORF\_ID:0231#4"

gene

CDS

complement(4977. 5393)  
/gene="csgr"  
complement(4977. 5393)  
/gene="csgr"  
/note="ORF\_ID:0231#6  
similar to SwissProt Accession Number P52104"  
/codon\_start=1  
/transl\_table=1  
/product="Curli production assembly/transport component  
CsgG precursor."  
/protein\_id="BAA35828.1"  
/db\_xref="GI:4062611"  
/translation="MRVHAVLMLISPLSMAGTMTFQFRNPFNGNPNNGAFLNS  
AQONSKYKDPYNDGFIETPSALDNTQAIOGILGLLSNINWTKGPRGMVNDYIV  
DIARNDQLOLVNDRKTQGTSTIOVSGLONNSYDF"  
complement(5418. 5807)  
/gene="csgr"  
complement(5418. 5807)  
/gene="csgr"  
/note="ORF\_ID:0231#7  
similar to SwissProt Accession Number P52105"  
/codon\_start=1  
/transl\_table=1  
/product="Curli production assembly/transport component  
CsgE precursor."  
/protein\_id="BAA35829.1"  
/db\_xref="GI:4062612"  
/translation="MKRYLRIYVAEFLFAAGNLHAEVEVPGLLTDHTVSSIGHDEY  
RAFSDKMSDVTGNLTINERPSARMSGITTYVQDVIPTQFLPLKRDKEKYVVAL  
IQREVALNRQINALLSTGDIADDER"  
complement(5812. 6462)  
/gene="csgr"  
complement(5812. 6462)  
/gene="csgr"  
/note="ORF\_ID:0231#8  
similar to SwissProt Accession Number P52106"  
/codon\_start=1  
/transl\_table=1  
/product="Probable csgAB operon transcriptional regulatory  
protein."  
/protein\_id="BAA35830.1"  
/db\_xref="GI:4062613"  
/translation="MNEVGHSTHCHTLLITKSSLOATALLQHKSLAITGKLIHQ  
RSLDDISSGSITLLDMEADSKLITHWQDTLSRNNNKKILLNLTPEDPYRDIENMP  
HINQVFSMEQDERVYVLOGVLRGECYFTQKLASYLITSGNRYNSTEALITTHEE  
KEILNKLRIAGSANNIARSLFISENTVTKTHLYNLFKKIAVKNRQAVSWANDNLNR"  
7217. 7672

gene

CDS

complement(4977. 5393)  
/gene="csgr"  
complement(4977. 5393)  
/gene="csgr"  
/note="ORF\_ID:0231#6  
similar to SwissProt Accession Number P52104"  
/codon\_start=1  
/transl\_table=1  
/product="Curli production assembly/transport component  
CsgG precursor."  
/protein\_id="BAA35828.1"  
/db\_xref="GI:4062611"  
/translation="MRVHAVLMLISPLSMAGTMTFQFRNPFNGNPNNGAFLNS  
AQONSKYKDPYNDGFIETPSALDNTQAIOGILGLLSNINWTKGPRGMVNDYIV  
DIARNDQLOLVNDRKTQGTSTIOVSGLONNSYDF"  
complement(5418. 5807)  
/gene="csgr"  
complement(5418. 5807)  
/gene="csgr"  
/note="ORF\_ID:0231#7  
similar to SwissProt Accession Number P52105"  
/codon\_start=1  
/transl\_table=1  
/product="Curli production assembly/transport component  
CsgE precursor."  
/protein\_id="BAA35829.1"  
/db\_xref="GI:4062612"  
/translation="MKRYLRIYVAEFLFAAGNLHAEVEVPGLLTDHTVSSIGHDEY  
RAFSDKMSDVTGNLTINERPSARMSGITTYVQDVIPTQFLPLKRDKEKYVVAL  
IQREVALNRQINALLSTGDIADDER"  
complement(5812. 6462)  
/gene="csgr"  
complement(5812. 6462)  
/gene="csgr"  
/note="ORF\_ID:0231#8  
similar to SwissProt Accession Number P52106"  
/codon\_start=1  
/transl\_table=1  
/product="Probable csgAB operon transcriptional regulatory  
protein."  
/protein\_id="BAA35830.1"  
/db\_xref="GI:4062613"  
/translation="MNEVGHSTHCHTLLITKSSLOATALLQHKSLAITGKLIHQ  
RSLDDISSGSITLLDMEADSKLITHWQDTLSRNNNKKILLNLTPEDPYRDIENMP  
HINQVFSMEQDERVYVLOGVLRGECYFTQKLASYLITSGNRYNSTEALITTHEE  
KEILNKLRIAGSANNIARSLFISENTVTKTHLYNLFKKIAVKNRQAVSWANDNLNR"  
7217. 7672

gene



CDS /gene="cs9B"  
7217..7672  
/note="ORF\_ID:0231#9"

## Alignment Scores:

Pred. No.: 7.98e-77 Length: 15047  
Score: 1115.50 Matches: 237  
Percent Similarity: 60.44% Conservative: 90  
Best Local Similarity: 43.81% Mismatches: 159  
Query Match: 41.128 Indels: 56  
DB: 1 Gaps: 9

US-10-066-551-4 (1-525) x D90741 (1-15047)

QY 33 SerCySerSer-----TrrleuProProleuGluGluArgThrGluSer--- 47  
DB 10775 TCCTGCATTCAGTACTGAGGCTTATGTTACCTC-GTGGCGAGCGCTGAATCGCAG 10833  
QY 48 ---ArgHisPheAsnThrSerLysProVal-----ArgLeuAspAsnIle 61  
DB 10834 TAAAAACGTTTCAGATTATTAACCCGTACCGCTTACCGCAGAGTACTTGTCT 10893  
QY 62 LeuGlnIleArgHisThrProHisThrAsn----- 71  
DB 10894 GTTATGATGAGAAAGCCGACCTCTACGAAAGACTCTTACCAACAAGAGATGAT 10953  
QY 72 -----GlyLeuSer 74  
DB 10954 GATTTCGCCGCGTGGCAGCGCGTCTGCCACTGTGTGCACAGATCCCGGTGATGT 11013  
QY 75 AspIleTyrLeuLeuAsnAspProHisGluAlaPheAlaIleArgAlaIleGlu 94  
DB 11014 GGCCTTTTCTCTGGAGAAAGTCTGATGCTTTCGCCGCGGTATGCTTGGCCGAA 11073  
QY 95 SerAlaGlnHisSerLeuAspLeuGlnIleTyrTyrIleTrrPargAsnAspIleSerGlyArg 114  
DB 11074 ATGGCAGACATACGCTCGATGTTCACTATTCATCTGGCAGCAGCATATGTCGGTGG 11133  
QY 115 LeuLeuPheAsnLeuValTyrLeuAlaIleGluArgGlyAlaArgValArgLeuLeu 134  
DB 11134 TTACTCTTTTCCGCTGTTAGCCGACGACAAACGCTGCGCTTCCGCTTGTGCTG 11193  
QY 135 AspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsn 154  
DB 11194 GAGGACACATACGCGCGACTTGACGACATTTACGCTTGGACATCTCCAGCG 11253  
QY 155 IleGluValArgLeuPheAsnProPheValLeuArgLysTrrPargAlaLeuGlyTyrLeu 174  
DB 11254 ATGAGAGTCGCGCTTTTAACTCTTCCGTTCCGCTTCCGCTTGGTATATC 11313  
QY 175 ThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArg 194  
DB 11314 ACCGACTTTTCCGCTTAACTGCGCGATGACAAATAAAGTTTCACTGCGATGGCGTG 11373  
QY 195 AlaThrIleLeuGlyAlaArgAsnIleGlyAspGluTyrPheLysValGlyGluAspThr 214  
DB 11374 GTGACCTGTGGAGGAGCAAAATATTGCGATTTTGGAGAGGAGAGCA 11433  
QY 215 ValPheAlaAspLeuAspIleLeuAlaThrGlySerValAlaGlyGluValSerHisAsp 234  
DB 11434 CTCTTTTCCGATTTAGATGTCATGCAATGAGACCCGCTGTAAAGACGCTTCCGATGAT 11493  
QY 235 PheAspArgTrrPrrAlaSerHisSerAlaHisAsnAlaThrArgLelleIleArgSerGly 254  
DB 11494 TTGGCCGCTACTGTGATTCGAATCGGTTTCACTTACAGAGGGCTG----- 11544  
QY 255 AsnIleGlyLysGly-----LeuGlnAlaLeuGlyTyrAsnAsp 267  
DB 11545 GATGTCCCGAGAGGTGAATGGCGATCGCATGAGTTACCGCGCTCTGCGCAATACGAT 11604  
QY 268 GlnThrSerArgHisAlaLeuLeuArgTyrArgGlnThrValGluGlnSerProLeuTyr 287

DB 11605 GCCATGACCAT-----CGTATTATACGCAAAATGAAATCCATTCATTAA 11652  
QY 288 GlnLysIleGlnThrGlyArgIleAspTrrPrrGlnSerValGlnThrArgLeuIleSerAsp 307  
DB 11653 AATCATCTGGTGTGATGACACATTGCGCTTATCTGGCGAGACGCTTATTATTAAGAT 11712  
QY 308 AspProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyLeuGln 327  
DB 11713 GATCCGCGAAGAGGAGGCGCAAGGCAAAAGGCAATTCATCTTACCGCAGCGCTGTTC 11772  
QY 328 AspAlaLeuLysGlnProGlyLysSerValTyrLeuValSerProTrrPheValProThr 347  
DB 11773 GATATCATGGCTGACCCACCATGACCATGATATATCTCTTCCTATTTTGTACCGACA 11832  
QY 348 LysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeu 367  
DB 11833 CGCGAGGTGTGGCGCACTTACGATGTGAGAAAGGGGTAAAGATTGCGATCTTA 11892  
QY 368 ThrAsnSerLeuGlnAlaThrAspValAlaIleValHisSerGlyTyrValLysTyrArg 387  
DB 11893 ACCAATTCCTTCCGCTAACGATGTTGCTGTGCTCATCGCGATACGCGCGCTGGCC 11952  
QY 388 LysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValPro 407  
DB 11953 AAAAAATTGCTCCGCTATGCGTGAATTTATGAACTCAAGCCGACGCGTGAACAAGT 12012  
QY 408 AlaThr-----LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThr 426  
DB 12013 AGTACGTTACACATCGCGCATACCGCATACCGGTAATTCGAGCCGACCGCTGCTAAACC 12072  
QY 427 PheIleValAspGlyLysAspGlyIlePheIleGlySerPheAsnLeuAspProArgSerAla 446  
DB 12073 TTTAGCATCATGATGTAACAGGATGTTATGCGCTCTTCAATTTCCATCCGCTTACACA 12132  
QY 447 ArgLeuAsnThrGluMetGlyValIleGluSerProLysIleAlaGluGlnMetGlu 466  
DB 12133 TTCTCATAATCTCAAAATGGGCTTCGTGATGAGACGCAAAACGCTGGCACGTTAATGAT 12192  
QY 467 ArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsn 486  
DB 12193 AAACGCTTATTAGACACGATGATGCGCGCTGGCAGCTCCGCTGGCAGGTGGGGA 12252  
QY 487 ArgLeuGlnTrrPrrHisAspProAlaThrArgLys-----ThrrTrrProAsnGluProGlu 504  
DB 12253 CGGATCAACGCGGTGATTCATTCATGCAAGAAAGAGATTTTTCGCAAAAAAGAACCCGCC 12312  
QY 505 AlaLysLeuTrrPrrLysArgIleAlaIleAlaLysIleLeuSerLeuProIleGluSerLeu 524  
DB 12313 ACCAGTTTCTGGAAGCGGTTATGTCAGACTGCGCATATTGCGCGGGAATGGTTA 12372  
QY 525 Leu 525  
DB 12373 TTG 12375  
RESULT 6  
D90742 19201 bp. DNA linear BCT 07-FEB-1999  
LOCUS Escherichia coli genomic DNA. (23.8 - 24.2 min).  
DEFINITION D90742 AB001340  
ACCESSION D90742.1 GI:1651513  
VERSION Complete and shotgun sequencing: cs9A; ycdE; cIs; nov; mdog; mdoh;  
KEYWORDS ycdE; msyB; ycdE; htrB; ycdE; ycdE; cybA; sox2; pycC.  
SOURCE Escherichia coli(strain:K12) DNA, clone:Kohara clone #232.  
ORGANISM Escherichia coli  
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE  
1 (sites)  
Oshima,T., Albe,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,  
Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,  
Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,  
Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nishimoto,H.,  
Nishio,Y., Saito,N., Sempel,G., Seki,T., Tagami,H., Takemoto,K.,  
Wada,C., Yamamoto,Y., Yano,M. and Horiiuchi,T.

TITLE A 718-kb DNA sequence of the Escherichia coli K-12 genome  
corresponding to the 12.7-28.0 min region on the linkage map

JOURNAL DNA Res. 3 (3), 137-155 (1996)

MEDLINE 97061202

REFERENCE 2 (sites)

AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kasimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y., Oshima, T., Salto, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.

TITLE The systematic sequencing of the Escherichia coli genome in Japan

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 19201)

AUTHORS Mori, H.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail: hmori@etc.aist-nara.ac.jp, Tel: 81-7437-2-5660, Fax: 81-7437-2-5669)

COMMENT Collaboration Information:  
Project:  
The Japan E. coli genome DNA sequencing project

Group:  
The Japan E. coli genome DNA sequencing group

Members: (1995.4 - 1996.3)  
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kasimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y., Oshima, T., Salto, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.

Headed by:  
Name: Takashi Horiiuchi  
Address: National Institute of Basic Biology, Okazaki, 444, Japan  
E-mail: kishori@nibb.ac.jp  
Information operator:  
Name: Hirotsada Mori  
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan  
E-mail: hmori@etc.aist-nara.ac.jp  
URL:  
The Japan E. coli genome database  
http://b3w3.aist-nara.ac.jp.  
Location/Qualifiers  
1. 19201

FEATURES  
source  
/organism="Escherichia coli"  
/strain="K12"  
/db\_xref="taxon:562"  
/map="23.8-24.2 min"  
/clone="Kohara clone #232"  
/note="Nucleotide position 1105083-1124283 from the initiation site of 'thra (0 min)'. This clone is from Kohara lambda miniset library."  
1. 186

gene  
/gene="csga"  
<1. 186

CDS  
/gene="csga"  
/note="ORF\_ID:0231#10  
similar to SwissProt Accession Number P28307"  
/codon\_start=1  
/transl\_table=1  
/product="Major curlin subunit precursor."  
/protein\_id="BA35840.1"  
/db\_xref="GI:1651514"  
/translation="DLTGKGFSGNSATLDWNGKNSKSEMTVKFGGNGAADVDTASNS  
VNVTVGFGNNAATHAY"  
245. 577  
/gene="ycde"  
/gene="ycde"

CDS  
245. 577  
/gene="ycde"  
/note="ORF\_ID:0232#1  
similar to SwissProt Accession Number P52107"  
/codon\_start=1  
/transl\_table=1  
/product="Putative curli production protein precursor (orf)."  
/protein\_id="BA35841.1"  
/db\_xref="GI:4062620"  
/translation="MNTLLILAAISSQITFMTQGGDYITLPEVTLTQSLCRVOLL  
SLRESSSQSQSKQKKTSLPANQPIALTKLSLNSPDRKIVTVSDGSLHSQ  
WPSSEKS  
2034. 2345  
/note="ORF\_ID:0232#2"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BA35842.1"  
/db\_xref="GI:4062621"  
/translation="WEPFLNSLMGSLFPPFLIATAGSTVGGVITHEYGQIVPACDV  
STQSSPEVMNCPOGNSIPGKTYSSKALMSGVKNQAQIASVXVYLDKQKLAIVNIEY  
N"  
2440. 2973  
/note="ORF\_ID:0232#3  
similar to PIR Accession Number I39569"  
/codon\_start=1  
/transl\_table=1  
/product="ORF2"  
/protein\_id="BA35843.1"  
/db\_xref="GI:4062622"  
/translation="MKRTIHVVGDTITKLAVDVIVNANPSLMGGGVGDAIIRAGP  
ALLDACLVRQGGCCPTGHAIVITLAGDLPKAVVHTGTPMRGGEQMDLDDATL  
NSLRIVANSTISVAFPAISTGVGTPRAAAAEIAVKTSEITRHALPEQVITFCYD  
EENAHLYERLLTQGD"  
2915. 4396  
/gene="cls"  
/note="nov"  
2915. 4396  
/gene="cls"  
/note="ORF\_ID:0232#4  
similar to PIR Accession Number A56145"  
/codon\_start=1  
/transl\_table=1  
/product="Novobiocin resistance-related protein Nov"  
/protein\_id="BA35844.1"  
/db\_xref="GI:4062623"  
/translation="MMKRTPTSTKDSLPNKEMDLPLASAVLPLCSQHPGCGIFPL  
EKSLDAFAARYRLAEMAHPTLDVQYIIMDDMSGILFSLAALAKRGVRYRLLDN  
NTPGDDTLRLDLSHPRIEVLFPNFSRLRPLIGYTDPSRLNRMHNSKFTVDGYV  
TLVGRNIGDAVFPAGGEPLFSDLDVMAIGVVEVDADDFARYCKSVSLQGYDV  
PEGEMADRIELPASWHDAMTHRTLRKSESPFTNHLVDGTLPLWAKTRLSDPAK  
GEGAKRHSLLPQRLFDIMGSPSEKIDLISSFPYPTAGVQQLRMVKGAKIALTN  
SLAANDVAVVHAGYARMRKLLRGVELEYELKPEQSGSTLHDGITGNSGASLHAKT  
FSIDKTVFISGFNDPDSSTLNTNEMGVISEETLAOLIDKRFIOSGDAMQRLDR  
MGRINVDVRAHAKKEILLKEPATSPKRWVRLASILPEVLL"  
complement(4404. 5561)  
/note="ORF\_ID:0232#5"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BA35845.1"  
/db\_xref="GI:4062624"  
/translation="MNPVPARQREYFELDSIRAWMLGLGPIHSLISYSHTWVHVAS  
SLMLFLFNDLTHSRMQYFVVISGFSTMLRLPLKMKWAVRVERGIPMLTAPLL  
TLPOFIMLOYKVAESWPGSLYDKNTLAMELISHMLVAVMTLCVWLFKRI  
RNNLENSDKTKKRSMYLSVIFLCIGYAVIRFTFIVPPLISNGMNFIVMOTL  
FYLPEFIIIGALAFIEPHLKALFTTPSRGCTLAALAFAYALLNORYSGSDAMYETES  
VITWVIGLMMVNVSEGRRLINFSQAVTYEVANALAFIYLVHHPDLTFEAYITPHI  
TSNMLGFLCGLIFVVGTAIIILYEHLRIPLKLPFSGKPVVKKENDAPAR"  
5935. 7490  
/gene="mdoc"  
5955. 7490  
/gene="mdoc"  
/note="ORF\_ID:0232#6

```

similar to PIR Accession Number S35417"
/codon_start=1
/transl_table=11
/product="mdog protein."
/protein_id="BA035846.1"
/db_xref="GI:1651515"
/translation="MMKRWLAAVMLTYTSSWAFSIDVAKOASLAKGKETPK
SNLPSVDFDMKYADVOOIOFNHDKAYNNLTKPELFEYHOGMEFDPVKINEVATA
VKRIKSPDYFDGVDHDKDVKDGLGAGFRVLVPIKSNKDEIVSMGLASVPRVI
GAGOVYGLSARGLAIDFALPDSGEERPREKREVIERTKRLTYIYALDSPRATG
KVVVPGRDIVDVOSKLYLNDKVGKLVGLVLTSLFEPNOSPANNRPELHDSNG
LSIHAGNEMIMRPLNPKHLAVSSFSNENQGGGLDRGDFSEFEDLDYDLPDS
AMVTPKEMGKSGELVEIPIINDENDIVAYWIPDQPEEGKEWFKYTTSEFSD
KLHAPDNAMVOQTRSTGDKOSNLIRPDQTLFVDFVFGAEMKLLPEDTPVTAOTS
IGDNGEIVESTVRNPNYTKGMRVWVYKXAKKTTEMRALVANADQTLSEWYQLP
ANF"
CDS
complement(6057..6314)
/note="ORF_ID:0232#7"
/codon_start=1
/transl_table=11
/protein_id="BA035847.1"
/db_xref="GI:4062625"
/translation="MLNTASEIIRAVFDSDCGSHFTYDRGIEVHTLWELEFEW
GLQIVPVRFVMIKLDLIVRFHIAENGGVAFGRVAFPG"
7483..10026
/gene="mdoh"
7483..10026
/gene="mdoh"
/feature="mdoh"
/feature="ORF_ID:0232#8
similar to PIR Accession Number S35418"
/codon_start=1
/transl_table=11
/product="mdoh protein."
/protein_id="BA035848.1"
/db_xref="GI:1651516"
/translation="MNKTEYIDAMPPIASKEALPKTDIAVHQAALDAERTWARED
DSPQSVARLEQAMPDSDLADGOLIKDEGRDOLKAMPEARKSMFPDPMRTNPVGRF
Alignment Scores:
Pred. No.: 1,096-76 Length: 19201
Score: 1115.50 Matches: 237
Percent Similarity: 60.448 Conservative: 90
Best Local Similarity: 43.818 Mismatches: 159
Query Match: 41.12% Indels: 56
Gaps: 9
US-10-066-551-4 (1-525) x D90742 (1-19201)
QY 33 SerCysSerSer-----TrrleuProProLeuGluGluArgThrGluSer--- 47
Db 2793 TCCTGCATCACTACTGGGGTTATAGGTACCCCTC-GTGGCGCAGCGGCTGAATCGCAG 2851
QY 48 ---ArgHisPheAsnThrSerIysProVal-----ArgLeuAspAsnIle 61
Db 2852 TAAAAACCGTTTCAGAAATTATATACCCGCTTACCCGACAGGATATACCTTTGTCT 2911
QY 62 LeuGlnIleArgHisThrProHisThrAsn----- 71
Db 2912 GTTATGATGAAAGAAACCGCCACCTCTACGAAAGACTCCTTACCCACAAGAGATGAT 2971
QY 72 -----GlyLeuSer 74
Db 2972 GATTGGCCCCGGGTGGCGAGCGGTGCTGCACATGTTGCCAACATCCCGTCACTG 3031
QY 75 AspIleTyrLeuLeuAsnAspProHisGluIlePheAlaIleArgAlaIleLeuIleGlu 94
Db 3032 GGCCTTTTTCCTGAGAAAGCTGTGAGTGGTTGGCCGCCGGATATGCTGTGGCGGAA 3091
QY 95 SerAlaGluHisSerLeuAspLeuGlnIleTyrTyrIleTyrPargAsnAspIleSerGlyArg 114
Db 3092 ATGGCGACACATACGTCATGTTCAGTATTCATCTGCGACAGACATATGTCGGTGG 3151
QY 115 LeuLeuPheAsnLeuValTyrLeuAlaIleGluArgGlyValIleArgValArgLeuLeu 134

```

```

Db 3152 TTACTGTTTCCGCCCTGTAGCCGACGAACGCGGTGGCTGCGTGTGTTGTCG 3211
QY 135 AspAspAsnAsnThrArgGlyLeuAspPheLeuAlaLeuAspSerHisProAsn 154
Db 3212 GACGCAACAAATACGCCCCGACTTGCAGCATTTATACGCTTGCAGTCATCCACGC 3271
QY 155 IleGluValArgLeuPheAsnProPheValLeuAspGlyTrrPargAlaLeuGlyTyrLeu 174
Db 3272 ATTGAATCGCGCTTTTATCTTTCCTGTTCCGTTGCTGCGTCCGCTGCTGATTATAC 3331
QY 175 ThrAspPhePargArgLeuAsnArgMetHisAsnIysSerPheThrAlaAspAsnArg 194
Db 3332 ACCGACTTTCCCGCTTATGCGCTATGCGCATGAATAAAGTTTACCTGCGTGGCGTG 3391
QY 195 AlaThrIleLeuGlyGlyArgAsnIleGlyAspGlyTyrPheIysValIleGlyIysPthr 214
Db 3392 GTACACCTGTGGAGAGCAAAATATGTGATGCTATTTGGACACGCGAGAGCA 3451
QY 215 ValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyIleValSerHisAsp 234
Db 3452 CTTTTCGGAATTTAGATCTCATGGCAATAGGACCGGTGTAGAGACGCTGCGATGAT 3511
QY 235 PheAspArgTyrTrrPalaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGly 254
Db 3512 TTCGCCCGCTACTGCTATTCGAAATCGTTTACCTTACAGACGAGTGGCTG----- 3562
QY 255 AsnIleGlyIysGly-----LeuGlnAlaLeuGlyTyrAsnAsp 267
Db 3563 GATGTCCCGGAGGTGAATAGCGGATCCGATCCGATCCGCTGCGATTAACGAT 3622
QY 268 GluThrSerArgHisAlaLeuLeuArgTyrArgGlyThrValGluGlnSerProLeuTyr 287
Db 3623 GCCATGACCAT-----CGTTATTTACGAAATGGAATCCGATCATTTATA 3670
QY 288 GlnIysIleGlnThrGlyArgIleAspTrrGlnSerValGlnThrArgLeuIleSerAsp 307
Db 3671 AATCATCTGTTGATGAAACATATGCGCTTATCTGGCGACAGACGTTTATTAAGTAT 3730
QY 308 AspProAlaIleGlyLeuAspArgAspArgGlySerProProIleAlaIleArgLeuGln 327
Db 3731 GATCCGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3790
QY 328 AspAlaLeuIysGlnProIleIysSerValTyrLeuValSerProTrrPheValProThr 347
Db 3791 GATATCATGGCGTCAACCCAGTGAACGATGATATATCTCTTCATTTTGTGACCGACA 3850
QY 348 LysSerGlyThrAspAlaLeuAlaIysLeuValGlnAspGlyIleAspValThrValLeu 367
Db 3851 CCGCGAGGTGTGGCGCACTCTTACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3910
QY 368 ThrAsnSerLeuGlnAlaThrAspValAlaIleValHisSerGlyTyrValIysTyrArg 387
Db 3911 ACCAATTCCTTCCGCTACCAATGTGTGTCGTCACGAGGAGGAGGAGGAGGAGGAGGAGG 3970
QY 388 LysProLeuLeuIysAlaGlyIleIysLeuTyrGlnLeuGlnProAsnHisAlaValPro 407
Db 3971 AAAAATTCCTCCGCTATGCGGTGAGATATATGAACTCAAGCGAGCGGTGAACAAAT 4030
QY 408 AlaThr---LysAspIysGlyLeuThrGlySerSerValThrSerLeuHisAlaIysPthr 426
Db 4031 AGTACCTTACAGAGATCGCGCATTAACCGGTAATTCGCGAGCGAGCGCTGATGTAACAC 4090
QY 427 PheIleValAspGlyIysArgIlePheIleGlySerPheAsnLeuAspProArgSerAla 446
Db 4091 TTATGACATGATGATTAACAGGATGTTATGCGCTTTCGTTAAATTCGATCGCGCTTACA 4150
QY 447 ArgLeuAsnThrGluMetGlyValIleGluSerProLysIleAlaGluGlnMetGlu 466
Db 4151 TTGCTCAATACGAAATGAGGCTTCGATAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 4210
QY 467 ArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsn 486

```

Db 4211 AACGCTTATTCAGACGCCATGATGCGGCTGCGACCTCCGCTGCGACAGGTGGGA 4270  
Qy 487 ArgLeuGlnTrpHisAspProAlaThrArgLys-----ThrTyrProAsnGluProGlu 504  
Db 4271 CGGATCACTGCGGTGATGCATGCAAGAAAGAGATATTCTCAAAAAGAACGCCGCC 4330  
Qy 505 AAlaLysLeuTrpLysArgIleAlaIleAlaLysIleLeuSerLeuLeuProIleGluSerLeu 524  
Db 4331 ACCAGTTCTCGAAGCGGTGTATGTCAGACACGCGGTCGATATTCGCCCGCGAATGTTA 4390  
Qy 525 Leu 525  
Db 4391 TTG 4393  
RESULT 7  
LOCUS AP002555 222605 bp DNA linear BCT 07-MAR-2001  
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 6/20.  
ACCESSION AP002555 BA000007  
VERSION AP002555.1 GI:13360886  
KEYWORDS  
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub\_strain:RIMD 0509952)  
DNA.  
ORGANISM Escherichia coli O157:H7  
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
REFERENCE  
AUTHORS 1 (sites)  
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,  
Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,  
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,  
Sasakawa,C. and Shinagawa,H.  
Complete nucleotide sequence of the prophage VT2-Sakai carrying the  
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7  
derived from the Sakai outbreak  
Genes Genet. Syst. 74 (5), 227-239 (1999)  
20198780  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS 2 (sites)  
Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,  
Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and  
Hayashi,T.  
Comparative analysis of the whole set of rRNA operons between an  
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an  
Escherichia coli K-12 strain MG1655  
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)  
20557356  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS 3 (sites)  
Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,  
Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,  
Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and  
Shinagawa,H.  
Complete nucleotide sequence of the prophage VT1-Sakai carrying the  
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli  
O157:H7 strain derived from the Sakai outbreak  
Gene 258 (1-2), 127-139 (2000)  
20564182  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS 4 (sites)  
Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,  
Yokoyama,K., Han,C.-G., Ohnsubo,E., Nakayama,K., Murata,T.,  
Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,  
Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and  
Shinagawa,H.  
Complete genome sequence of enterohemorrhagic Escherichia coli  
O157:H7 and genomic comparison with a laboratory strain K-12  
DNA Res. 8 (1), 11-22 (2001)  
21156231  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS 5 (bases 1 to 222605)  
Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and  
Hayashi,T.  
Direct Submission  
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome  
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,  
Japan (E-mail:ken@gen-info.osaka-u.ac.jp/  
URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,

COMMENT  
FEATURES  
source  
Fax:81-6-6879-2047)  
genome project.  
Location/Qualifiers  
1..222605  
/organism="Escherichia coli O157:H7"  
/strain="O157:H7"  
/sub\_strain="RIMD 0509952"  
/db\_xref="taxon:8334"  
74..607  
/gene="Ecs1423"  
74..607  
/gene="Ecs1423"  
/gene="Ecs1423"  
/note="Similar to B1045\_ECOLI g111787283 percent identity  
100 in 177 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative polypeptide"  
/protein\_id="BAB34846.1"  
/db\_xref="GI:13360887"  
/translation="MKTRIHVQGDITKLAADVIVNANPSLMGGGVGDAIHRACG  
ALIDACKLRQOGDCPTGHAVITLADLPAAKAVHVGVPVARGGEONEDLQDALV  
NSLRIVANSTYSVAPPAISGVGYRRAAAEIVATVSEFIRHALLPQGVYVCD  
EENHIXERILITQOGDE"  
549..2030  
/gene="Ecs1424"  
549..2030  
/gene="Ecs1424"  
/note="Similar to YMDG\_ECOLI g111787284 percent identity  
99 in 493 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative synthase"  
/protein\_id="BAB34847.1"  
/db\_xref="GI:13360888"  
/translation="MMKTRPSTDSIPNKEMNDLPPLASNVLPICSQHGGCGLPFL  
EKSIDAFRARTRLAEMSEHTLDVQYIIMODMSRLTFSALAAKGVNRRLDDN  
NTPGLDILRLDLSHPRIEVLNFPSEFRLRLPGYITDSRLRRNRNRSFYVDGV  
TLVGRNIGDAVFGAGEPLFSDLDVAIGVVEDVADDFARVYCKSVSPLOQVLY  
PEGEMADRIELPASMHNDAMTHRYLRKMSSEPTINHLVDGTLPLIMKTRLSDDPAK  
GEGAKRHHSLPQRLFDPMGSPSERIDIISSFVPTRGVAAQLIRVARKVKATILN  
SLANDVAHVSYAGYARWRKILRGVGLIELKPTREOSSTLHDGRTGNSGASIAHT  
FSDGKIVFTGSENFDPSTLNTNTEKGFVIESETLAQIDKRFIOSQYDAMQRLDR  
WGRINWDRHAKKKEIVLKEPATISFMRVWRILASILPEWLL"  
complement(2038..3195)  
/gene="Ecs1425"  
complement(2038..3195)  
/note="Similar to MDQC\_ECOLI g1130252265p175920 percent  
identity 99 in 385 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="glucans biosynthesis protein"  
/protein\_id="BAB34848.1"  
/db\_xref="GI:13360889"  
/translation="MNPVPAPQREVELDSIRAMLMGLIPHSILYSHTWVNSAP  
SLMTLLNDFHSFRMOVFVYISGYFYMFLRPLKRWKRVREVERGIPMLTAIPLL  
TLPEFIMLOYKGAESMPGSLDYDKTAMELISLMELVLVWTTCTGATVEFKI  
RNNLENSDKTKRKFSMWKLSVIFCLGIGVAAVIRRTFIIYPTLISGTAETES  
FYLPFLGLALAFIPPHUKALFTTPSGCTLAALAVAILNDRSGGAMVETES  
VITWVGLMNVNVFSFGRHLNQSARVTFVNASLFTLVHHPITLTFGAVITPHI  
TSMWGLFCGLIFVAGIAILYELHRIPLKFLFSKPVVKKRENDAPAR"  
3588..5123  
/gene="Ecs1426"  
3588..5123  
/gene="Ecs1426"  
/note="Similar to MDQG\_ECOLI g111787286 percent identity  
99 in 511 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11

```

/product="periplasmic glucans biosynthesis protein"
/protein_id="BAB34849.1"
/db_xref="GI:13360890"
/translation="MMKRMRLSAVMLTLTYSSMAISIDVAKOASLACKYEAPK
SNLPSYERDMKYADYOOIOPNHDKAYNNNLTPKLEPHOGKMFDPKINERTATA
VKRIKSPDYFTFGDVHDKDTPKDLGPKRVLYPLNSKNEIYVSLASFRVLA
GAGVYGLSANGLAIDTALPGSEEPREFKEIMRPKPTDRLTYALALDPRATGAY
KFVVPGRDLYDVQSKLYLRDKYGLGVLATSMLEFGPQSPANNYRELDSDNG
LSIHAGNGEMIRPLNNPKHLAVSFSNGFGLORGFDFRELDLDRYDLRPS
AWVPKGEKGSVEIYEIPIPNDETNNIYVMTPDLPBEGKEMNPKRYTTTFSRDE
KLAPPDNAVQOTRRSTGDVAKOSNLIRPDGCTATFVDFTEAKMKLPEDTPVATIS
IGDNGEIVESTVRNPTKGMRLVVRVKKDAKKTTEMRALVNADOTLSETWSYOLP
ANE"
/gene="7659"
/gene="ECs1427"
5086. 7659
/gene="ECs1427"
5086. 7659
/gene="ECs1427"
/note="similar to membrane glycosyltransferase MDOH_ECOLI
g11787287 percent identity 100 in 847 aa but differs at
N-ter (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="membrane glycosyltransferase"
/protein_id="BAB34850.1"
/db_xref="GI:13360891"
/translation="MKPGATSYLPNNKTTXYIDAMPIAASEKALPKTDIRAVHQAID
AEHTMARREDSPOGSVKARLEOAMPDSALIKDEGDOLKAMEARSSMFPD
PMRTNPGYRFMDRLRGNDVTPRYTLRKEQSEKQKRTYGTTRITLYLLTAAQY
VATVMTKTLPGQWALINPDWVGODLWVSFQOLPLPYMLDTGILLLEAVLFCWSAG-
FMTALMGFLQLLIGRDKYSISISVGDPELBEHRTALIMPICNEDEVNRVAGARATW
ESVATKANAKHFDDYILSDSYNPDIQVAEQKAMMELIAEVGEGQIEYRRRRRRK
SGNIDFCRMRGSSQSYXVVLADSVMTGDCIGVRLMEANPNAGIISQSPKASGM
TLVRCOOFATRVYGPLFTAGLHFMOGESHYMGHNALIRKPREHLALPDEGGS
PAGSIIHSDPVEALMRAGCWITAYDLPSYSELPIINDLDELKDRDRKHGHNMF
RLFLVKGHPYRAVFLTGWSTYLSAPLMFLSTALQVYHATLEPYTLQPROLF
PVPQMRBELIALFASTVMLLEPLKLSILLIMCKGKEVGEWRVLLSLLEVLFS
VLLAPRMLFTLTVFVSAFLGMEVWNSPORDDSTSGEAFKRHSQLLGLWAVG
MAMDLRFLFLIAPLIVSLISPEVSYSSRATGLRTKRMKLELIPESYSPQVLAD
TDRLENNRORSLDDGFMAHVNPSFNALAMADRARASKVLEIADREVEOLNE
TPRELNRDRLYLLSDPYTMARLHFRWNSPERYSWVSVEGIKINLARKDPAAS
Q"
/gene="8059"
7832. 8059
/gene="ECs1428"
7832. 8059
/gene="ECs1428"
/note="identical to YCEK_ECOLI g11787288 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB34851.1"
/db_xref="GI:13360892"
/translation="MRLIVSIVMTLLSGCSISRTIPGGHGNQYPGVQWDVRS
AMRVVTLIDLPFSVFDLPLPIDIHGPE"
complement(8060. 8437)
/gene="ECs1429"
complement(8060. 8437)
/gene="ECs1429"
complement(8060. 8437)
/note="similar to MSYB_ECOLI g11787289 percent identity
99 in 125 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="acidic protein MSYB"
/protein_id="BAB34852.1"
/db_xref="GI:13360893"
/translation="MMTMYATLEAIDAREEFLADNPGIDAEDNAVQOFPNAOKYVLQ
DGDIMQVEFFADSGESECPLMLSGEAGASVPDGYDEIETROMQMENTLHMEDES

```

Alignment Scores:  
Pred. No.:

3.74e-75

Length:

222605

```

Score: 1113.50 Matches: 236
Percent Similarity: 60.44% Conservative: 91
Best Local Similarity: 43.62% Mismatches: 159
Query Match: 41.04% Indels: 56
DB: 1 Gaps: 9
US-10-066-551-4 (1-525) x AP002555 (1-222605)
QY 33 SerGysSerSer-----TyrLeuProProLeuGluGluArgThrGluSer 47
427 TCTTCGATCAGTACTGAGGCTTTATGCTTACCTC-GTGGGACGCGCTGAATCGCAG 485
48 ---ArgHisPheAsnThrSerLysProVal-----ArgLeuAspAsnIle 61
486 TAAAAACCGTTTCAGATTTATTACCCGTCAGCTTTACCCGACAGCATATTCTGTCT 545
62 LeuGluIleArgHisThrProHisThrAsn----- 71
546 GTTATGATGAAAGAAACCGCCACTTACGAAAGACTCCTTACCCACAGAGATGAT 605
72 -----GlyLeuSer 74
606 GATTTCGCCCGCTGCGGACGCCGCTGCTCCACACTGTGTCGCAACATCCGCTAGCT 665
75 AspIleTyrLeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaLeuIleGlu 94
666 GGTCTTTTCCCTCTGGAGAAAGTCTGATGGTGTGGCGCGGATAGCTGTGGCCGGA 725
95 SerAlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrArgAsnAspIleSerGlyArg 114
726 ATGTACAGACAAATACGCTGATGTTCAGTATTCATCTGACGACAGCATATGTCGGGTCGG 785
115 LeuLeuPheAsnLeuValTyrLeuAlaAlaGluValArgValArgLeuLeuLeu 134
786 TTACTGTTTCCGCGCTGTACCGCGACAAACGCTGGCTGCGCTGTTGTTGCTG 845
135 AspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsn 154
846 GACGACAAACAAATACCGCCGACTTGACGACATTTTACGCTGCTGACAGCCATCCCCC 905
155 IlegIuValArgLeuPheAsnProPheValIleArgGlyTyrPArgAlaLeuGlyTyrLeu 174
906 ATTGAAGTCGCGCTTTTAACTCTTTCGTTGCTGCTGCTGCTGCTGCTGCTATATC 965
175 ThrAspPheProArgLeuAsnAspArgGluMetHisAsnLysSerPheThrAlaAspArg 194
966 ACCGACTTTTCCCGTCTCATGTGCGCTATGACACAAATAAAGTTTCACTGTGATGGCGTG 1025
195 AlaThrIleLeuGlyGlyArgAsnIleGlyAspGluTyrPheLysValGlyGluAspThr 214
1026 GTGACGCTGTGTAGGGGGGCGGAATATGTGTATGCTTATTTGGGCGAGGGAGAGCCA 1085
215 ValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAsp 234
1086 CTTTTCGATTTTAACTATGTCATGCAATAGGACCCGCTAGTAGAGACCTTCGCCGATGAT 1145
235 PheAspArgTyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleLeuArgSerGly 254
1146 TTCGCCCGCTACTGCTGATTCGAATCGATTTCACCTTACAGCAGAGTGGTG----- 1196
255 AsnIleGlyLysGly-----LeuGluAlaLeuGlyTyrTrpAsnAsp 267
1197 GATGTCGCCGAGAGTGAAATGGCGGATGCCATGCAATGACCGGCTCTGCGCATACGAT 1256
268 GluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGluInserProLeuTyr 287
1257 GCCATGACCAT-----CGTTACTTACGCAAAATGGAATGCAATCCGCTTATATA 1304
288 GluLysIleGluThrGlyValArgIleAspTrpGlnSerValGluThrArgGluInserAsp 307
1305 AATCATCTGCTGTTGATGAACGCTTTCGCTTATCTGGGCGCAACACGCTTATTAAGTANT 1364
308 AspProAlaLysGlyLeuAspArgAspArgGlySerProProIleAlaGlyArgLeuGln 327

```

```

Db 1365 GATCCGGCGAAGGAGGCGCAAGCAACGCAATTCCTGTACCGCAGCGCTGTTTC
OY 338 ApsalaeLysglnProglulysSerValtyrLeuValSerProtyrPheValProthr
Db 1425 GATATCAGGGCTCACCCAGTACGATCATATTATCTCTCCATTTTGTACCGACA 1484
OY 348 LysSerGlyThrApsalaeLysValGlnAspGlyIleAspValThrValLeu 367
Db 1485 CGCGGAGGTGGCGCAACTCTTACGAGTGTGAGCAAAAGGGGTAAAGATTGCGATCTTA 1544
OY 368 ThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValIleTyrArg 387
Db 1545 ACCATTCTCTTGGCGGTACGATGTGTGTGTCGTCATCGCCGATACCGCGCGTGGCGC 1604
OY 388 LysProLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValPro 407
Db 1605 AAAAATTGCTCCGCTATGGCGTGAATATATGAACCAACCGCAGCGGTACAAAGT 1664
OY 408 AlaThr---LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaYsthr 426
Db 1665 AGTACGTTACACAGATCGCGCATTAACCGGTATTCGGTCCAGCGCTCATCTTAAAC 1724
OY 427 PheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAla 446
Db 1725 TTTACATCGATGATGTAACGGTGTATTACGCTTCTTCAATTTGATCCGCTTCAACA 1784
OY 447 ArgLeuAsnThrGluMetGlyValIleGluSerProLysIleAlaGluGlnMetGlu 466
Db 1785 TTGCATATACGAATGGGCTTGTGATAGACGCAAAACGCTGCTCATTAATTTGAT 1844
OY 467 ArgThrIleAlaAspThrSerProLutyrAlaTyrArgValThrLeuAspArgHisAsn 486
Db 1845 AAACCTTTATTACAGACCAAGTATATGCGCTGGACGCTCGTGTGACAGGGGGGA 1904
OY 487 ArgLeuGlnTyrPheHisAspProAlaThrArgLys-----ThyTyrProAsnGluProGlu 504
Db 1905 CGGATCACTGCGGTGATGTCGTATGCAAAAGAAAGATTTGTTCTCAAAAAGAACCCGT 1964
OY 505 AlaLysLeuTyrPheArgIleAlaAlaLysIleLeuSerLeuProIleGluSerLeu 524
Db 1965 ACCACTTTCGAAAGCGGTTATGTCAGACTGGCGTGAATTTGCCGTGGAATGTTA 2024
OY 525 Leu 525
Db 2025 TTG 2027

RESULT 8
AE006196/c 9255 bp DNA linear BCT 08-MAR-2001
LOCUS Pasteurella multocida pm70 section 163 of 204 of the complete
DEFINITION genome.
ACCESSION AE006196 AE004439
VERSION AE006196.1 GI:12721978
KEYWORDS
SOURCE Pasteurella multocida.
ORGANISM Pasteurella multocida.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Pasteurella.
REFERENCE 1 (bases 1 to 9255)
AUTHORS May,B.J., Zhang,Q., Li,L.L., Paustian,M.L., Whittam,T.S. and
Kapur,V.
TITLE Complete genomic sequence of Pasteurella multocida, pm70
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (6), 3460-3465 (2001)
MEDLINE 11248100
PUBMED 2 (bases 1 to 9255)
REFERENCE Zhang,Q. and Kapur,V.
AUTHORS Direct Submission
TITLE Submitted (24-OCT-2000) Department of Veterinary Pathobiology,
JOURNAL University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN
55108, USA
FEATURES Location/Qualifiers

```

```

source
1. .9255
/organism="Pasteurella multocida"
/strain="pm70"
/db_xref="taxon:747"
/complement(95. .940)
/gene="rpoH"
/note="PM1584"
/complement(95. .940)
/gene="rpoH"
/codon_start=1
/transl_table=11
/product="rpoH"
/protein_id="AAK03668.1"
/db_xref="GI:12721979"
/translation="MNKDTQTMVLVPOGSIETAYIRAANEYPMLEAEKELEAEPLYO
EDLEAAKILISHLEFVYIHVARGSYGLPQADLIQEGNIGIMAKVAFNEVEGVRY
SEFAVHAKIEIHSEHLEVRNMRIVKATTAQRKLEFNNRKTORIGMNEVEVANE
LGVSADVMESESMTGADVAFDLPDQNETTSPSYLDEKGSNFAALEENVEYEO
ATEQLSVALEGLDERSLDITIKARWLDEKATLQDLAKYINISAEVRQLETNALKLK
SAVNF"
/complement(1171. .2577)
/gene="PM1585"
/complement(1171. .2577)
/gene="PM1585"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03669.1"
/db_xref="GI:12721980"
/translation="MKYLOITITWKSRELFECVSIIPALIMTLTISDASPTKYONY
LELLEPAISMTVTAIPFYSYIIVGFVYVWKSSEAYDIOYLQWNSDVITFLG
GFLAEARKTKLDLLEKAVLPKFGNPNVVLGLMLVTAINMSLNSDATTAMTA
TISPLFRLNESSMLSKALLGIPAAASVGGMTIISAPNAIVAGLEKLGPISTL
EMMVGTPPLAILLIFEMRVLYKRYEINKESLDFSELOETPAHSNRVEFLKTIYL
VILAVTLFCMLMSKMGIPVNAASGIPVGLTIGVLESEKIDLPMDTILMLVAGGA
LGIAEIORIRAHVEOISHOVSEFIMLSLFGFTVILSNFMSMTAATITLIPMGVS
LLAWGCMNDIHPMLTIPVIGSASCALFPLPSTPMAIARSTGLIKOSERLGSVITIG
FLGPILSMLTALISPL"
/complement(2760. .4304)
/gene="PM1586"
/complement(2760. .4304)
/gene="PM1586"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK03670.1"
/db_xref="GI:12721981"
/translation="MLIVILIKLEFLALILFSLVILALISYORLPTSDRPISTHPL
NPNGLARHILPOVAOHPNLTGLVPLDGDKAQFLARLASEHAHTLIDQIYIMHNTV
SGHLLIOSIKYKAAARGVYKRLLDNNTKGMDTLASLNHPNIOIRLFPNMOROYR
WLGFLSPFFRLNRHMKSPFADCVNSILGGRNIGDEYFDVGNCVLEFADLDVARTGAV
VTHIQTFDRKWNSSPYPLESIIIRPITFPNLPALDETOYILQLDELPAKSL
KAGTLAFTWAEALISDDPKALKSLIDSVLNAIHAPTMNAANNLIIVSPVPH
VGVDFLSRISOTGQVSIILNLEATDVSIVHSGYAKHRTLOLOKQIOLKELPHAT
OMESSGHLKIGASSASLHAKFTLTDNRYLFGVSPNMDPRSMLEMTENGILLIDSEPLA
ILSDGLQONQANYAFSVKLNDAQLYWETOENCKWIKITENDPHTSWKRKRSVCLSNL
PVERHL"
/complement(4493. .5452)
/gene="PM1587"
/note="PM1587"
/complement(4493. .5452)
/gene="PM1587"
/codon_start=1
/transl_table=11
/product="Teha"
/protein_id="AAK03671.1"
/db_xref="GI:12721982"
/translation="MIRQSKPPPLPINTYFCIIIGLSALSLAMRYATPCEGATVSE
SLRAVSAMVLAFIGAAYAKWFSHQDLADLIDHLEKGNVSLIPITTLIGIGLVY
NMLEGTVLISGLAIQIAFAFYIIPAMRKIHQAEMLTIPAIYATVATNFSATLGVY
LGSELMGVLPAGFESMVLVLEPIVLOERLTLPDVEPLPMIGIQLAPAFVCGSAVL
TLNGKRLDELPAIMLIGYGVLLFLFLPLMIFVNGPTPAMMAFSCFLASMAKVALLE
YQNSDNLATLTLALFIFANFSIGMLNLNTGYLVYKGRFFIK"

```

```

gene      complement(5449. 5760)
          /gene="PM1588"
CDS       complement(5449. 5760)
          /gene="PM1588"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAK03672.1"
          /db_xref="GI:12721983"
          /translation="MNFNTLNTLPRIEDYVGLAIMDGETOIHITPATAGKLSLRV
          NALAEAFHGRIDRTSAQGIQLFAHSEDAKOPKHPHIDLLFQITHDLYQLPQ
          K"
gene      complement(5773. 6798)
          /gene="murB"
          /note="PM1589"
CDS       complement(5773. 6798)
          /gene="murB"
          /codon_start=1
          /transl_table=11
          /product="murB"
          /protein_id="AAK03673.1"
          /db_xref="GI:12721984"
          /translation="MNIQPHHTPSLPYQAQKIITTDIEQLKQMAACQTKKLVLL
          LGQSNVLFKDFQGVLLNRLMGIOHTQDADFHLHVNNGENHQLVAMSIOQIGF
          LBNLALPGCAGSAPLQNIAGVDFKDVCEVLELHNTNEIFRLSBOCFGRDS
          IFKHAYATGYVIAVGLKSKMIFVLKYSLANEKSAVNAAQIIFDEVCALRSKLP
          DPAEFGNAGSEFKNVSVKDPEDHLOEYPDIPHPDAGRVKLAAGLIDCGCKGY
          OLGGAHVHQAQVLYLNQDATTSSDVVELAHIRKVAQRFAMVQPEVRFIDEGEV
          DSEQALIC"
gene      complement(6802. 7425)
          /gene="PM1590"
CDS       complement(6802. 7425)
          /gene="PM1590"
          /codon_start=1
          /transl_table=11
          /product="unknown"
          /protein_id="AAK03674.1"
          /db_xref="GI:12721985"
          /translation="MKLFLEASFADVDFLEDPAGNIONKSIETPTASKVEEKY
          VDDARNALQOGAFIDELDTASLAETDRAKMEIYVSGNTEPFLLOELTRGTA
          OLIOOHIOAGKRIECSAGAMISAPNIDYKYMDSLDAQPMRLDRTALNDVYVYPH
          YTNPEVBAQKIIDYTSSTLPLOIDNHQALVNNETKIIKGNHKG"
gene      complement(7422. 9128)
          /gene="narQ"
          /note="PM1591"
CDS       complement(7422. 9128)
          /gene="narQ"
          /codon_start=1
          /transl_table=11
          /product="NarQ"
          /protein_id="AAK03675.1"
          /db_xref="GI:12721986"
          /translation="MKIKHSVSTKIANLIETITMTRAGVITSLSIMTMSNKSDASIN
          VSGSLRMOYRLHELEHPELVETGLKQYRISLSPALLEIDOMFVPEVQSYHOL
          IKRMERMYAKQREINAYKNEIADYVAQVDFQVFLPFAFKKNIIVALLISLMLS
          IIAVAYVYVYARKVEVPLKQIASIOVQOFKSHYVDFQKDEGLNLRARFSOM
          SESELEKIVLSSESVNEKTOKROTNRSLTMVYESSOQLTTNVESTIOHVKNIET
          SEHLRYMALVGEAHMNIREGOKAONDCCQVTLQIDEEKLAVLYMAGFPCCPRT
          MRNIOIISRTLYFKTKOROOQQLLMERARIIAELHDSLAQVIAFLQIQTLLKHN
          INOKSEEAQOSINITIASFEALRDGYVOLRELLATFRLTVEANKDKLAEVVDLSR
          NOTDIOMVOCSLPAQSFNAKQVHILQIVRESVNAIKHSNGRIEVAIHANDEGEY
          EFTVDRDDIGIPSLDEPRGHYGLNIMHERATQNLNHLRIANRPSGTEIKTLHQRN
          LT"
BASE COUNT 2719 a 1883 c 1953 g 2700 t
ORIGIN
Alignment Scores:
Pred. No.: 2,15e-76 Length: 9255
Score: 1106.50 Matches: 238
Percent Similarity: 61.99% Conservative: 80
Best Local Similarity: 46.39% Mismatches: 170
Query Match: 40.79% Indels: 25
DB: 1 Gaps: 8

```

```

US-10-066-551-4 (1-525) x AE006196 (1-9255)
QY 23 LeuIIleSerLeuLeuCysLeuLeuLeuCysSerCysSerSerTrpLeuProProLeuGlu 42
    ||| |||||::: ||| ||| ||| |||||
Db 4256 CGTTTAAAGCCTGTGATACTTGCCTTAATAATGATACCAACGA---TTACC----- 4209
QY 43 GluArgThGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeu 62
    ||| ||| ||| |||||::: ||| ||| ||| |||||
Db 4208 -----ACATCCGATGCTCCATCAGTACGATCCCTCCCTTATATCCCAAGGGCGTTA 4155
QY 63 -----GluIleArgHisIsthProHisIsthAsnGlyLeuSerAspIle 76
    |||||::: |||||::: |||||::: |||||
Db 4154 GCACGCCATATGCTTGCCGCAAGTTGCCCAACATCCCACTTAACGGGATG----- 4104
QY 77 TyrLeuLeuAsnAspProHisGluIlePheIleAlaIleArgIleAlaIleGluSerAla 96
    ||| ||| ||| |||||::: ||||| ||||| ||||| |||
Db 4103 TATCCTTTAGGCGATGGGAAGATGCTTTCGCTGCGCTTGCGCTGACGCAACATGCC 4044
QY 97 GluHisSerLeuAspLeuGlnTyrTyrIleTrpArgAsnAspIleSerGlyArgLeuLeu 116
    |||||::: |||||::: |||||::: |||||::: |||||
Db 4043 GAACATACACTGCACCTTACAGTTATTTATTTGSCATACAGATGTGTGTCACCTTACTC 3584
QY 117 PheAsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAspAsp 136
    ::||| ||||| |||||::: |||||::: |||||::: |||||
Db 3983 TTCAAAAGCTATATATAAGCCGACGCGCGAGTAAAGTCCGTTATATTATGGACGAC 3924
QY 137 AsnAspThrArgGlyLeuAspAspLeuLeuAlaIleAspSerHisProAsnIleGlu 156
    |||||::: |||||::: |||||::: |||||::: |||||
Db 3923 AATTAATACCAAGCATGACGACGATGCTTGCTTCACTCACTCAACACCTTAATATCCAA 3864
QY 157 ValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp 176
    ::||| |||||::: |||||::: |||||::: |||||::: |||||
Db 3863 ATCCGCTTATTCATCCCTTATGCAACGTCATATCGTTGGTAGCTTTTGACGAC 3804
QY 177 PheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThr 196
    ||| ||||| |||||::: |||||::: |||||::: |||||::: |||||
Db 3803 TTTTATTAAGCTCAATGCTGCTATGCAACAATAATCCCTTCCACCCAGATGCGTCATAGCT 3744
QY 197 IleLeuGlyGlyLysArgAsnIleGlyAspGlyTyrPheLysValGlyGluAspThrValPhe 216
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||
Db 3743 ATTTCTGGCGGGGTATATATGTGATGAATATTTGATGTGGTAAAGTGCTGATATTATTT 3684
QY 217 AlaAspLeuAspIleLeuAlaThrGlySerValValGlyValSerHisAspPheAsp 236
    |||||::: |||||::: |||||::: |||||::: |||||::: |||||
Db 3683 GCTGATCTTGATTTGCTCCGCAACCGCGCGCTGTTACACACTTCAGATGATTTTGAT 3624
QY 237 ArgTyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIle 256
    |||||::: ||| |||||::: |||||::: |||||::: |||||
Db 3623 CGCTATTGGAAATGACCATCAAGTTATCTTTGAAAAGCATTTATCATACGATCTCATTT 3564
QY 257 GlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGlyIsthSerArgHisAlaLeuLeuArg 276
    |||||::: |||||::: |||||::: |||||::: |||||
Db 3563 ACGCCTTTTAATCCGTTGCTGCTGATTAAGATGAAGAAACCCAA-----ACT 3519
QY 277 TyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrArgLysArgIleAsp 296
    ||| ::| ::| ::| ||| |||||::: |||||::: |||||::: |||||
Db 3518 TATTTAAACAACACTACTGACGCTTTCGTAAGCTTAAGAGGACCACTTACGCT 3459
QY 297 TrpGluSerValIsthArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAsp 316
    ::||| |||||::: |||||::: |||||::: |||||::: |||||
Db 3458 TTTACATGGCAGCAGCAGCATTAATACAGCATGATCTTAAAGAGCAGCTAGCAATCA 3399
QY 317 ArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSer 336
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 3398 TTGATTCACAGACAGTGTCTTAGCACAATATTGCTCCACAAATGTAATAGCGAAACAC 3339
QY 337 ValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLys 356
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 3338 TTAATCATTTGTGCGCTTATTTTGTACCAACCATGTTGTGTGATTTCTTAAGTCGT 3279
QY 357 LeuValGlnAspGlyLysAspValThrValLeuThrAsnSerLeuGlnAlaThrAspVal 376
    ::| ||| ||| |||||::: |||||::: |||||::: |||||::: |||||

```

```

Db 3278 ATCAGCACACTGGCACACAGGTTCCATTCTAACCAATTCTACTGAGACACCGATGTT 3219
Qy 377 AAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLys 396
Db 3218 AGCATTTGATCATCTGCTGTTATGCTTAACACCGCAAAACATTTGTTACAAAGCAATCCAA 3159
Qy 397 LeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly-----Leu 414
Db 3158 CTTGATGACCTGAACCT---CATGCACACATTCAATGAATGAAACCCACAGGCATTATTA 3102
Qy 415 ThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIle 434
Db 3101 AAGAGGACCGAGTAGCCCGCATTTACATGCCAAACATTTACTCTTATATGCTATTTA 3042
Qy 435 PheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyVal 454
Db 3041 TTTGTTGGTTCCTTAAATGATGATCCGCTCGCTGCTATGTTGAATACGAAATGGGATTA 2982
Qy 455 ValIleGluSerProLysIleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro 474
Db 2981 CTCATTGACAGCCCTGAATGCTGCTTACTGATGATGATGACGACAAATCAAGA 2922
Qy 475 GluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAla 494
Db 2921 AATTAATGCTTTTTCGCTTAAGTTAATATGAGCGCCACAGCATGTATTGGCAACGCAAGA 2862
Qy 495 ThrArgLys-----ThrTyrProAsnGluProGluAlaLysLeuThrLysArgIleAla 512
Db 2861 AATGGAAATGATTCGCTATGAATGAATGAGCACACACAGGTTGCTTAAGCGTTTACT 2802
Qy 513 AlaLysIleLeuSerLeuLeuProIleGluSerLeuLeu 525
Db 2801 GTGTGCTGTTTACTGTTGTTACCTGTTGAACATTGCTTA 2763

RESULT 9
AE005315 10190 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
DEFINITION of 155
ACCESSION AE005315 AE005174
VERSION AE005315.1 GI:12514572
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria: Proteobacteria; gamma subdivision: Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 10190)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10190)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimantanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE Location/Qualifiers
1. 10190
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"

```

```

gene
/note="enterohemorrhagic"
617..1072
/gene="cs9B"
/note="21675"
617..1072
/gene="cs9B"
/function="Structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 151 of 151 are 100.00 pct identical
to residues 1 to 151 of 151 from Escherichia coli K-12
strain MG1655: B1041"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar to CsgA"
/protein_id="AAG55787.1"
/db_xref="GI:12514573"
/translation="MKNLKLMMLTILGAPGIAAGYDLANSENFVNLKRSFN
QALIGQNTNNSQLRGSGSKLLAVNDESSNRKIDQNGDNLVATPDAGSANDA
SISQATYNTMIITIOKSGSKANITTYOTKTAIVYVQROSMATRYTOR"
1113..1571
/gene="cs9A"
/note="21676"
1113..1571
/gene="cs9A"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 152 of 152 are 96.71 pct identical to
residues 1 to 151 of 151 from Escherichia coli K-12 strain
MG1655: B1042"
/codon_start=1
/transl_table=11
/product="curlin major subunit, coiled surface structures;
cryptic"
/protein_id="AAG55788.1"
/db_xref="GI:12514574"
/translation="MKLKVAIAIAYVSSGALAGVYPOYGGGNGHGGNSGPN
SLDNYOYGGNSALALQADARNSLDTTHQGGCAGADYGGSDSDSLDTRGGNSA
TLDDNGKDSIMTYKQFGGNGAAYDQIASNSTYVTVYVGRNNATRAHQ"
1630..1962
/gene="cs9C"
/note="21677"
1630..1962
/gene="cs9C"
/function="putative structure; Not classified"
/note="Residues 1 to 110 of 110 are 98.18 pct identical to
residues 1 to 110 of 110 from Escherichia coli K-12 strain
MG1655: B1043"
/codon_start=1
/transl_table=11
/product="putative curli production protein"
/protein_id="AAG55789.1"
/db_xref="GI:12514575"
/translation="MNALLLAALSSQITFTTQGGWYTIPEVTLTQSLCRVOIL
SLREGSSQSQTQKEKTLSPANQPIALTKLSINSPDRKIVTVVSDGSLHSQQ
WPSSEKRS"
2083..2394
/gene="21678"
2083..2394
/gene="21678"
/function="Orf; Unknown function"
/note="Residues 1 to 103 of 103 are 98.05 pct identical to
residues 1 to 103 of 103 from Escherichia coli K-12 strain
MG1655: B1044"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55790.1"
/db_xref="GI:12514576"
/translation="MFRPLNSLMGSLFFPIAIASTAGGVIHYHGYQVACDY
STQSSPYEMNCPQNGSVPGKTYSSKALMSGVNNAOIASVVOYLDRKILAVNIEX
N"
2489..3022
/gene="21679"

```





```

Db 3501 CTTTTCGATTATGATGTCATGCAATAGAGCCCGTGTAGAGACGTTGCCGATGAT 3560
Oy 235 pheaspargTyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleLeuSerGly 254
Db 3561 TTCGCCCGCTACTGCTATGCAAAATGCGTTTCCCTTACAGACGTTGCTGCTGCTG 3611
Oy 255 AsnIleGlyLysGly-----LeuGlnAlaLeuGlyTyrAsnAsp 267
Db 3612 GATGTCGCCGGAAGGTGAATAAGCGGATCCGATCCGCTCTGCGCAATACGAT 3671
Oy 268 GluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyr 287
Db 3672 GCCATGACCAT-----CCTTACTACGCAAAATGCAATCCAGTCCGTTTATA 3719
Oy 288 GlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAsp 307
Db 3720 AATCATCTGTTGATGACGTTGCCGCTTACTGCGCAAAACAGCTTATTAATGAT 3779
Oy 308 AspProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyLeuGln 327
Db 3780 GATCCGCGCAAAAGGAGGAGGCAAGCAACGCAATTCATCTGACCGCAGCCCTGTTTC 3839
Oy 328 AspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheValProThr 347
Db 3840 GATATCATGGGCTCACCCAGTACGATGATATATCTCTTCTTATTTTGTACCGACA 3899
Oy 348 LysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeu 367
Db 3900 CCGCGAGGTGGCGCATCTTACGATGTGAGCAAAAGCGTAAGAATGCGATCTTA 3959
Oy 368 ThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArg 387
Db 3960 ACCAATTTCTTCCCTTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4019
Oy 388 LysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValPro 407
Db 4020 AAAAATTTCTTCCGCTATGCGTGAATTAATGAATCAACGCGCTGACCAAACT 4079
Oy 408 AlaThr---LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThr 426
Db 4080 ACTAGCTTACACGATCGCGCATACCGGTAATTCGCGTCCAGCCGCTGCTAAACG 4139
Oy 427 PheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAla 446
Db 4140 TTTACATCGATGTAACGCGTGTATCGTCTTCTTCAATTTGATCCGCGTCAACA 4199
Oy 447 ArgLeuAsnThrGluMetGlyValAlaIleGlnSerProLysIleAlaGluGlnMetGlu 466
Db 4200 TTGCTCATACTGAATGGCTTCTGATAGAGAGGAAACGCTGCTCAGTTAATGAT 4259
Oy 467 ArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsn 486
Db 4260 AAACGCTTATTCAGACGATGATGCGGCTGCGCTGCGCTGCGTGCAGAGTGGGGA 4319
Oy 487 ArgLeuGlnThrHisAspProAlaThrArgLys-----ThrTyrProAsnGluProGlu 504
Db 4320 CGGATCACTGGGTGATGCTGATGCAAGAAAGAGATTGTTCTCAAAAAGAACCCGCT 4379
Oy 505 AlaLysLeuTrpLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGlnSerLeu 524
Db 4380 ACCAATTTCTGAAACGCGTGTATGCTAGACACTGCGCTGATATTTGCCGTGAATGTTA 4439
Oy 525 Leu 525
Db 4440 TTG 4442

```

```

RESULT 10
LOCUS AE008749 22411 bp DNA linear BCT 31-JUL-2002
DEFINITION Salmonella typhimurium LT2, section 53 of 220 of the complete
genome.
ACCESSION AE008749 AE006468

```

```

VERSION AE008749.1 GI:16419641
KEYWORDS Salmonella typhimurium LT2.
SOURCE Salmonella typhimurium LT2
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE 1 (bases 1 to 22411)
McCllelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Portolillo, S., All, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22411)
The Salmonella typhimurium Genome Sequencing Project.
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs:
GLIMMER: http://www.tigr.org/softlab/glimmer/glimmer.html and
GenMark: http://opal.biology.gatech.edu/GenMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes: http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc:
http://ecocyc.org/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
Reguondb;
http://kitch.cifn.unam.mx:8850/db/reguondb\_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES
source
1..22411
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
434..1308
/gene="phoH"
/note="STM1126"
434..439
/gene="phoH"
/note="putative RBS for phoH; Reguondb:STMS1H001398"
454..1308
/gene="phoH"
/note="similar to E. coli phoB-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blast hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=1
/product="PhoB-dependent ATP-binding pho regulon
component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCAHRYQENAGDVYIOLKVLMSGROKAVIKARREKRVLRDS
RSIKORESEVSTSLVOMGVYALIGMARDSDRTSYKARNEQATLVAIDSKILFAT

```





```

/codon_start=1
/transl_table=1
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="SPTREMBL:Q82770"
/transl_table="MTQOPQAKYRHRDAPDYQITDIEFLDAEKTVAISQAVR
HGADAPARLIDGEDLIVSHVNDAPWPAVKEEGALLISDLPERFTLRIYNEISPA
NTALEGLVQSGDALCTOCSEAGFRHTIYDRPDVLAFTTKIADSKYFELNSG
RVAGLENGHVMQWODPEPKPCYLFALVAGDPDVLADFTTRSGREVALIELYDRG
NLDRAPAMTSLKSKMKNDETRGLIEDLDIYMTAVDFNMGAENKGLIFNSKY
LARTDTADKDYLDIERVYGHETPHNNTGNVTCRDMFOLSKBGLVFRQEESSDL
GSRVNRISNVRIRGLQFADASPMAPIRPDVIEMNFTYTLTYEKGAEVIRMIH
TLLEENRQKCMQLYFERHDSAACTODFOAMEDASNVDSLHFRMYQSQGTPIVY
KDDNPETEQYTLTISQTPATDOAEKOPHPIPAELIDNEGVNIPLOKGGHVNA
VLANTQAEQTFPDNVYEQPVALLCEFSAPVLEKMSDQOLFPLMHARNDSRMD
AASLANTYIKLANARHQQGPLSLPVHADAFAVILDEKIDPALAEITLPSANE
IAELFEVIDPITAIQVREALTRTILAELEDFLAIYNNHLDERTVDGDIKRTLRN
ACLRFLTFGETELANTLVSKOYRDANNMTDALAASAANAQLPCRDILMOEYDKWH
ODGLVMDKMFLLQSTSPAVENYLETVRSLSKHSFMSNPNRIISLIGFASGNPAFH
AODSGYQFIVEMETLIDNSRNDQVASRLIEPLIRLRYDKRQEKRAALBQLKLEN
LSDLYERITKALA"
532. 1332
/misc_feature
/gene="STY1078"
/feature="Pfam match to entry PF01433 Peptidase_M1, Peptidase
family M1, score 245.10, E-value 1.9e-84"
1060. 1089
/gene="STY1078"
/feature="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3001. 4011
/gene="STY1079"
/feature="pyrd"
3001. 4011
/gene="STY1079"
/feature="Orthologue of E. coli PYRD (PYRD_ECOLI); Fasta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="SPTREMBL:Q827S9"
/transl_table="MYPPFKALFQLDPERAHEFTFOQLRITGTPLEALVROKVP
KPVCMGLTFNPNGLAGLDKDECDALGAMGSGLEICIVPRPQGNDKPLFR
LVDAEGLINRGFNGLGVNDLVENVYKKAHFDGIGINIGKNKDTPEVNGKDYLCME
KVYAVAGYIAINISNPNGRLTLOYGDALDDLTATKKNKNDLOVHNKRVPAVKI
APDICEELIOVADSLRHNTDVIATNTTDRSLVQCKMCOOTGSGSLPDLKST
EITRRLSELKGLPIGVGIDSVIAREKTAAGATLVQITSGFTFGPLKEIVT
HI"
3136. 3990
/misc_feature
/gene="STY1079"
/feature="Pfam match to entry PF01180 DHODHase,
dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238. 3297
/gene="STY1079"
/feature="PS00911 Dihydroorotate dehydrogenase signature 1"
3874. 3936
/feature="PS00912 Dihydroorotate dehydrogenase signature 2"
4174. 4719
/gene="STY1079"
/feature="STY1080"
4174. 4719
/gene="STY1080"
/feature="Orthologue of E. coli YCBW_ECOLI; Fasta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"

```

---

```

/db_xref="SPTREMBL:Q827S8"
/transl_table="MMRIKPDNMRMYVDEEDHRMMLDLANGMLPRSRSEKMLTPDA
FCPTGEVDADALYFSFEKCKDPELTGEORREIVNLVAIRYKLPQMPKSMHVAH
GEMPRPGDASASVLSPTAROVNLVYPERGNALCILAQPGVIAGRTHQDGAIK
IMNDRKAPQVICHSESLQAV"
/feature="complement(4716, .5825)"
/gene="STY1081"
/feature="complement(4716, .5825)"
/feature="STY1081"
/feature="Orthologue of E. coli P75863; Fasta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="SPTREMBL:Q827S7"
/transl_table="WATLSRLFIHPVKSMRIGLTHALADISGLAFDRIFMYTESDGT
FITAQPFQWNRFTSPPLHDGLHLPAPGSSALVFTFTPDAPTEVGNHFTARVA
PTAIQMLSGFEFSRDVQLGWGPOLTBRVKRHNNAVPLFADGYPLLINASLRDQ
KCPACVOMEORFNPNIYVSGVAAEEDSKVLRIGVIDVYKPCSRCLFTYVSPKGO
KHPGSEPLATIQAFRTAODNGDVGDSGVLIRNSGVIRVGEVELIARAPAKVATG
LDSVTPPKHHDGASTYIDMOCOTFCGNNQOVLBDELEMGIRIRIPISCAIGCCCRIR
LLEGVSPLKKSAMGDDGTILSCSCVPXTALLEN"
/feature="complement(4737, .4955)"
/gene="STY1081"
/feature="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924. 8032
/gene="STY1082"
5924. 8032
/gene="STY1082"
/feature="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SH:YCBY_HAETN (P44524; P43945) fasta
scores: E(): 0, 60.3% id in 710 aa
orthologue of E. coli YCBY_ECOLI; Fasta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="SPTREMBL:Q827S6"
/transl_table="MNSLFASTARSLDELEKLGAVGCOVVGVRHGGDTRLI
YOSLMSLASLIIIPGCEKTYSLDLYFGVQALNMEIFNPGATFVFGSLGNDTI
RNSOYGMKGVADAIYDAFTKRLPNPNVDRSDPDIRVWMLNKETAISTALDSGGCL
LRGYDRGRGLAPIKETLAAIYMRSGKQPTFLDPKSGSTGLILEAMNATDRAPGL
HRGHWGSGMAQDEDAIWOEVYAEKQTRARKGLAYSSHFTGSDSDARYTERASNAR
RAGIGELITFEVKDYAQLSNLPKQPGYGVISNPPYGERLSDPALIALHSLIGRTMK
NOFGCMNLSLSEASPDLSGLQLRADKQFKAKNGPLDCVQKNYHIAETTADSKPATVA
EDYANRLRNKLKLEKMAROBIECTRYLVDADLPYNAVDPRYGMAYIOEYAPKPT
DAOKRORLPDIATLTVLIGIPPKVILKTPREKGNOKVQKSEKEELFVSYNMA
RLKVVULTYLDLDTGLFLDHRIRARMIGEMSKGDELNLNLSYTSASVYHAGLGARNTT
VDMSTRYILEMARNLRLNGLSGRAHRLIQADCLGLWRANDEFDILFTDPPFSKSR
MEESFDVORDHVALMKDLRLRKGTITMFSNNKGRFANDEGLAELGLTADQETIOKT
LSPDFARNQIHCWLIRAA"
5927. 7060
/misc_feature
/gene="STY1082"
/feature="Pfam match to entry PF01170 UPF0020,
uncharacterized protein family UPF0020, score 761.30,
E-value 3.7e-231"
6506. 6541
/gene="STY1082"
/feature="PS01261 uncharacterized protein family UPF0020
signature"
8045. 9952
/gene="STY1083"
8045. 9952
/gene="STY1083"
/feature="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap"
Fasta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa

```



TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Comparison of the genomes of two Xanthomonas pathogens with differing host specificities	Nature 417 (6887), 459-463 (2002)	22022145	12024217	2 (bases 1 to 10979)	Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melanda, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, Y.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and KitaJima, J.P.
Direct Submission	Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil				
TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS
Location/Qualifiers					
1. 10979					
/organism="Xanthomonas axonopodis pv. citri str. 306"					
/strain="306"					
/db_xref="taxon:190486"					
/note="pathovar: citri"					
complement(157..1719)					
/gene="cls"					
/note="XAC2871"					
complement(157..1719)					
/gene="cls"					
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="cardiolipin synthase"					
/protein_id="AA037716.1"					
/db_xref="GI:21109170"					
/translation="MTRKRFKVRGIALVILVALSLSYGGRFADRGQVSHYLPATAVPTDKVAPLQAHADDTGVILPDNDATRAVRLTRAKGRSLDLOYIWHADFTGNLHNEELLRADRGVRLLLDDNNIHGSDVLAALDHPLEIRLFNPTRAREGLTMGVGEVLMFESINRRMKNKAIADGRAPVAGVDEYFPAARTNMDAAIMGPVGOAEQVDFVWNSPMLPLSALVTAPOALEALRSLDGMASVRHPYVERLKRSPSVRELMOGDRPVMSQARIVSDPPEKAEQAPQADMTPIYLIGEMHHADRELKVISPPYFPGDEGMWIGHILRQNRVRSILNLSLANDVVAHSSYAGRVYLDLGVRLHELPKGRKPDGSLFGSSGASLHTKAFVDDSSCFISFNLDPRSMINTMGLFNDRVVVAELERLYNHRVSAVSYRTLDGELKMHDDAOPPEVMSREPAASVWRGGAAYMGWLPESQI"					
complement(1945..4047)					
/gene="XAC2872"					
complement(1945..4047)					
/gene="XAC2872"					
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="metallopeptidase"					
/protein_id="AA037717.1"					
/db_xref="GI:21109171"					
/translation="MKLRKLAPLSLAIVVALSACGASDDAKVAPAPAVPTKVDVSKLAPVIAFSSDLDPSIACCOILINGVNKKKANVPDRTSMGSEFYLARSLTIOHAILOEGLARSGISAGSVAGLADIMDTGSDDEAIDKAGITPLOPOLKIDALITLDAIAMLNRSVNOGGIVFSEBANDYKNSSEDMIAVAGGGIGLEBEKGYTPPAOKIRIOUYATRRMLELSGTIPPAQAQAKAWAETRLNANSLRIELDRDARKTNPDVAGANAITPHEFDMAFPALEVPAGTFESLSQPEYFELDAMLADTVPDVKARLREHTVDEAPYLAKEPEQANFDEHAKTLRGQDMLPMWKRTLNANVAMGEALCOLVQGSFPAESKEOMOLVONLSAALKARLOKIDMMSAETKORALEKASFTPKIGYPDQRMMSGIOTRGDGFLANQAAQAFNYRYMLDKIGKPVDRKEMHMPOTVNAVYNATTRNEIYEPALIOPEFDPKADPALNTNGCAVIGHEMHGXDDSGQFPAKGMDSMTDADKLETTERTQVLAQFDGYEAIKGVYHKKLTIGENIGDGGTLVATDIALQALQDEPAAKQIDGSDQDQREFENWMAVWRNFTDELRLNTDPHAPANFRANAPSMPAFQAQFCRPGDAMVKGDERVTW"					
complement(4324..4689)					
/gene="XAC2873"					
complement(4324..4689)					
/gene="XAC2873"					
/note="putative; ORF located using Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="conserved hypothetical protein"					
/protein_id="AA037718.1"					
/db_xref="GI:21109172"					
/translation="WVNSGAVRRCALRRVPRADHVHTQTWNPPADDAVYCAVLTVLGILVAGGALAGDISGAGLALLICVGIILLGFGVAGPSROMEMWADAAAGHEVALVLMVLAVPYVLMAPEYARR"					
complement(4773..5522)					
/gene="gpm"					
/note="XAC2874"					
complement(4773..5522)					
/gene="gpm"					
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="phosphoglycerate mutase"					
/protein_id="AA037719.1"					
/db_xref="GI:21109173"					
/translation="MTRKLVLLRHGOSQNLNDRFTGVNDELTDGROEAVAAGRLMKDEGLQDFVAHTSVLKRAHTHGTALKEIDQMLPYKSMRLNERHYGGLOGIDAETAAKHGEQVKTIRRSYDIPPPAMDYNDPRHPCHDRKRYATLDNRLPGTSLATTVLRYLPYHDAIAPQLKAGOTVLTVAHGSUNRLVYLVNDVSNQIETLELPIGILPELFDLSLQVRSERYLDGPEAKRAAEAVANOGKAK"					
5625..6338					
/gene="nfi"					
/note="XAC2875"					
5625..6338					
/gene="nfi"					
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="endonuclease V"					
/protein_id="AA037720.1"					
/db_xref="GI:21109174"					
/translation="MOTNDPVAGDGSYIQAROLQRLARRVLDADAASATPOLLAGFVGGEDDOCTRAAVALDDQTLPLETHIRVPTSMYVPGILSFRLEPLLOLAILSRPDLVDFIDGGIAHPRKGTIAHREGVVGLICIGIAKORLGSFAPRPPRDLATPILIGSSIGMALSKSPKPCNPLIYSPGRVSMOGALDPTLTLAYRLPEPTRLADRLASRGEVTPPGVYSGHL"					
complement(6409..7077)					
/gene="XAC2876"					
complement(6409..7077)					
/gene="XAC2876"					
/note="putative; ORF located using Glimmer/Genemark"					
/codon_start=1					
/transl_table=11					
/product="hypothetical protein"					
/protein_id="AA037721.1"					
/db_xref="GI:21109175"					
/translation="MIPRSIGMOVIAAKFQHERSCIASVPAOSKCMWPSYCHRAMNNGGPTTHGCRIGKRSRRRVGAHLASRSRSHVLDABEITLIPNLKNVYHGLDLRKNOIALLVMAASVPTLGTQSVQHAVSSRIISVAPGVRSILAFSANCKTVACAYV					







```

/db_xref="GI:21113885"
/translation="MPTLSEPMSTSTAITTEDFGRFRORFIFPAAGISISSGKKAMLCG
RLGRLREHODNNTQYQLLESRODRIOTAIIDLITNETYFFRPPKHLRLRA
DHRGGLPLRCWSSASTGEAYSMAMVLDLQAGFAVGTGDISNAPKARQYH
ALORIDGLIPQAYLKRYCLRGSEAYATLIVLERDRKVFVANAALPMGSPDVI
FLRWMYENCGOTKREVLIRVLANKPGHFCIGHSLSLSDIDLIOVAPSIFFKA"
4838..5896
/gene="cheb"
/ncbi="XCC2705"
4838..5896
/gene="cheb"
/ncbi="XCC2705"
/translation="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="protein-glutamate methyltransferase"
/protein_id="AAM41977.1"
/db_xref="GI:21113886"
/translation="MVVDSAVYRQVIVGLVNEAGIEVIATADPLAIEMKQOMP
DYIVLIDVEMPRMDGITFRLKIMSEPTVITCSTLTERGARTMDALAAGAVVTKP
RLGLQFLTDSADLVATVSAARANVRKLARVAPLEAVKHTADVILPAQGRRA
LAQTERIVAIGTSTGTQALEVLTALPRVCPGIVIQHMEKFTAAFAARLNLQO
IAVKAANNDRVMPGRALIAFGKHLRLRSGAGYFEVLEGPVNRHRSPVDVFRS
AARAGSNALGIIMTGMGDGAAGLEMRQAGARTIADDEHNSIVFGMPKKAIRKGA
DRILPLIGMAREIVTQLO"
complement(6196..7758)
/gene="c1s"
/ncbi="XCC2706"
complement(6196..7758)
/translation="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="cardiolipin synthase"
/protein_id="AAM41978.1"
/db_xref="GI:21113887"
/translation="MTKRIKFMAGIGVILVLAISVLSYXGGRADREGRASHALP
VAAQTPIDQVYAPLQAHACKTGMVILSDINDAFAVALTARAAGSLDIOYIWAH
DFTGLNLNELLRADRGVRRVRLLDNDNTHGSDVLAALDSHPLEIRLEPRTARE
GTLMKGVMLVLRMSINRRMNRKAMIADGRIAAVVGVRGDEYFPAARDTNDMDAA
LIGTPVROSEQVDFAYVNSPALPLAALVPTAKPOLAEOLRGLDAGMASVGHAYVER
LROSPVHALMGKGNRPVHMOQARIVADPPEKAEQAPOPDMPTILIGEMANAREL
KLISPEVFGDDGMKIGLRKRDVRSVLTNSIAANDVAVHSGYADYRIPILOQVY
RLHLEKPMGRPDGSLFGSSGASLHTKAPVVDNSAGFTSPNIDPRAMNLNEMGLFE
DRVTAELERLYHNRKVASPVSTRVLTLEGGRLMDDAQPRAVMTREPPASVWRGA
KVLGWLPLDSQL"
7854..8147
/gene="XCC2707"
7854..8147
/gene="XCC2707"
/translation="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="isoxac transposase"
/protein_id="AAM41979.1"
/db_xref="GI:21113888"
/translation="MVDSVMSKRYTDEKIEAVROVTDGRGFVAEVARLGVTH
SLVAMLRFTKSGVHRAEVDSAEVRLKAELRVTREEDILKKAAYFAKG"
8165..9019
/gene="XCC2708"
8165..9019
/gene="XCC2708"
/translation="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11

```

Alignment Scores:  
 Pred. No.: 8.77e-57  
 Score: 856.50  
 Percent Similarity: 53.58%

```

Best Local similarity: 39.06% Mismatches: 187
Query Match: 31.57% Indels: 60
DB: 1 Gaps: 15
US-10-066-551-4 (1-525) x AE012383 (1-10211)
QY 32 CysSerCysSerSerTrieuPro-----ProLeu 41
DB 7718 TCGTGTGCTGGCGCGTGGCAGCGCTGCTGCGCTGATGAGCTACGGCGCTTTGCCGATC 7659
QY 42 GIUGLuarGthGluSerarGhisPheasnthSerLysProValArg----- 57
DB 7658 GC-GACCGCGCGCCCTCCAGCCAT-----GCCCTCCGCGTGGCGCGCAGCGCAACG 7609
QY 58 ---LeuAsnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleu 76
DB 7608 CCAATGACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7552
QY 77 TyrLeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleuSerAla 96
DB 7551 GTGATCTCTCGCACAACTCATCGATCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCG 7492
QY 97 GIUHisSerLeuAspIleuInTyTyrTleTPArgAsnAspIleSerGlyArgLeuLeu 116
DB 7491 GCGCGGACCTGATCTGCACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 7432
QY 117 PheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAspAsp 136
DB 7431 CACAAAGCATTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 7372
QY 137 AsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGlu 156
DB 7371 ATGAACATCATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 7312
QY 157 ValArgLeuPheAsnProPheValLeuArgLys-----TPrArgAlaLeuGlyTyr 173
DB 7311 ATCCGCGCTTTCACACCCACCCGTCGACCCGACCCGACCCGACCCGACCCGACCCGACCCGACCC 7252
QY 174 LeuThrAspPheProArgLeuAsnAspArgArgMetHisAsnLysSerPheThrAlaAspAsn 193
DB 7251 GTTCGCGCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATG 7192
QY 194 ArgAlaThrIleGluGlyIleArgGlnIleGlyAspGlyTyrPheLysValGlyGluAsp 213
DB 7191 CGATCGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7132
QY 214 ThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyLysValSerHis 233
DB 7131 ACCAATTATGAGACATGACGACGCGGCTGATCGCGCGGACGCGGCTGATCGCGGACGACG 7072
QY 234 AspPheAspArgTyrTPrAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSer 253
DB 7071 GTTTCGATGCTTACTGGAACAGCCCAATGCTTCCCTGCGCGCGCTGCTGCTGCTGCTGCTGCTG 7012
QY 254 -----GlyAsnIleGlyLysGlyLeuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleuGln 264
DB 7011 AAGCGGACGCGCTGAGCACTGCGCGGACGCTGAGACCGCGGAGTGGCGTGGGCGGCG 6952
QY 265 TyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSer 284
DB 6951 GCGCACCGC-----TACGTGACCGCGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCG 6919
QY 285 ProLeuTyrGlnLysIleGlnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleuGlnIleu 302
DB 6918 CCAAGCGTCATGCTTTCATGCGAGGCAATCGCCGCGTCAATTTGGGTGAG---CAGGCG 6862
QY 303 ArgGluIleSerAspAspProAlaLysGly-----LeuAspArgAspArgArg 318
DB 6861 CCGATCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6802
QY 319 LysProIleIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyr 338
DB 6801 AGCGCGATCTGATGCGCGAGATGCGCAAGCA-----CGCGCGCAACTGAAA 6754

```

[illegible]

```

Oy 358 ValGlnaspGlyIleAspValThrValLeuthraSnSerLeuGlnAlaThrAspValAla 377
    ::::::::::::::::::::
Db 29362 GCTAAATGGGTTGAGTGGTATTTGACCAATTCATTGATGCCCGATGTTGT 29303
    ::::::::::::::::::::
Oy 378 AlaValHisSerGlyTyrValValGlyTyrArgLysProLeuLeuAlaGlyIleLysLeu 397
    ::::::::::::::::::::
Db 29302 ATTGTCATGTCAGGTTTCCACACTGGCGAAGCACTACTGTGGCGGCGATCTT 29243
    ::::::::::::::::::::
Oy 398 TyrGluLeuGlnProAsn---HisAlaValProAlaThrLysAspLys-----GlyLeu 414
    ::::::::::::::::::::
Db 29242 TTTGAGATTAATGATAGTCTCCAGATATCAAGATTAACGAAATCGTTTGGCGTACA 29183
    ::::::::::::::::::::
Oy 415 ThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIle 434
    ::::::::::::::::::::
Db 29182 CGCCAGCAGACGACCAAGTCTACAGCCGCAAGCGCTTGGCGTATGATGATCAATAA 29123
    ::::::::::::::::::::
Oy 435 PheIleGlySerPheAsnLeuAspProArgSerAlaArgLysAsnThrGluMetGlyVal 454
    ::::::::::::::::::::
Db 29122 TTTATTTGTTCTATATATATCTCCAGCCAGCTCAGCCCAATATCAATACCGAGCTGTGTT 29063
    ::::::::::::::::::::
Oy 455 ValIleGlySerProLysIleAlaGluMetGluArgThrLeuAlaAspThrSer--- 473
    ::::::::::::::::::::
Db 29062 CTTATCAAAAGATAGTAATTTGGCAGGACGCTACACAAAGCTTTGATATAGCCAGCC 29003
    ::::::::::::::::::::
Oy 474 ---ProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAsp 492
    ::::::::::::::::::::
Db 29002 ATCACCCTCAACGCTATGAACTTAACACTTACGCGCAAGGTAATATTAATGCGCAT--- 28946
    ::::::::::::::::::::
Oy 493 ProAlaThrArgLysThrTyrProAsn-----GluProGlu 504
    ::::::::::::::::::::
Db 28945 -----ACCATGCAAAATGTGTCAACACGATCATTTTGCATCAGACACCACAC 28901
    ::::::::::::::::::::
Oy 505 AlaLysLeuTrpLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGluSerLeu 524
    ::::::::::::::::::::
Db 28900 ATGAATCATTAAGATCGATTATTCATTGCGTGGCGACGACGATGCGATTCATTCGCTA 28841
    ::::::::::::::::::::
Oy 525 Leu 525
    ::::::::::::::::::::
Db 28840 CTT 28838
    ::::::::::::::::::::

RESULT 15
AP003017 208315 bp DNA circular BCT 15-MAY-2001
LOCUS Mesorhizobium loti plasmid pMLb DNA, complete genome.
DEFINITION AP003017
ACCESSION AP003017
VERSION AP003017.1 GI:11995001
KEYWORDS
SOURCE Mesorhizobium loti (strain:MAFF303099) plasmid:pMLb DNA.
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
AUTHORS Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
Watanabe,A., Iidesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpou,S., Sugimoto,M.,
Takeuchi,C., Yamada,M. and Tabata,S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
JOURNAL DNA Res. 7 (6), 331-338 (2000)
MEDLINE 21082930
REFERENCE 2 (bases 1 to 208315)
AUTHORS Kaneko,T.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
INSTITUTE, The First Laboratory for Plant Gene Research, Yana
152-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
LOCATION/Qualifiers
1. .208315
/organism="Mesorhizobium loti"

FEATURES
Source
/strain="MAFF303099"
/db_xref="taxon:381"
/plasmid="pMLb"
/complement(378..686)
/gene="ml19502"
/complement(378..686)
/gene="ml19502"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB54717.1"
/db_xref="GI:14028125"
/translation="MSRIDLDFEWRVYIQPLPNKPGVPRVDDRLVNGITFVYLRSG
ICFSPLYKARNLVEREFNKAKOPRIATRYDKLAENYLAALKLVSTRIMLRSESTP"
/complement(782..1144)
/gene="ml19503"
/complement(782..1144)
/gene="ml19503"
/codon_start=1
/transl_table=1
/product="transposase"
/protein_id="BAB54718.1"
/db_xref="GI:14028126"
/translation="MSRIDLDFEWRVYIQPLPNKPGVPRVDDRLVNGITFVYLRSG
APWRDLPERVGPRTCYNRFVRRKAGVMDRLMADITKAHDGEVOMIDTSIVRHQOG
ATARRGVEIILSVLGAA"
1273..2133
/gene="ml19505"
1273..2133
/note="ml19505"
/gene="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB54719.1"
/db_xref="GI:14028127"
/translation="MIRSSSSVFLLCLWLNSIIILASCSRPQPVSDADIDFLASDA
HVVGDVPLVLFIALTGTYAAKPSFLNRQDQQAATDRIEAFKKAINFOAPKID
OLEIAVAYAGWMDLDPKPFKRICARLSNOMSKSYCDPDMALAKQAMPENRYVVDK
FDADNYFTVGSQDLSHMQLVNEASVCDRNLSPSKYATICTAIIPIKHLIAV
WTVDSTGETHSRQARERGSITAFVHAIIQPSDFALLADVCKLNNPSSISPGSS
PCGSPISKSR"
/complement(2688..2891)
/gene="ms19506"
/complement(2688..2891)
/gene="ms19506"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB54720.1"
/db_xref="GI:14028128"
/translation="MTRKISDLGPPIPGKRHGAEPARVEDNRYPCSYCGQLVDQDLRQ
VGIRNWIPIKTRQSLCSRAKIAT"
/complement(3117..3422)
/gene="ml19507"
/complement(3117..3422)
/gene="ml19507"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB54721.1"
/db_xref="GI:14028129"
/translation="MRQMSRAPKVGDEITLTATILVMEGVSSVSLPSYNEPLAVD
SPAKAPGOKIEIRGFATRVDEKGVTVIRIDGGGLTVADSDITNVSAISRVSRO"
/complement(3482..4006)
/gene="ml19508"
/complement(3482..4006)
/gene="ml19508"
/note="unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB54722.1"
/db_xref="GI:14028130"

```



OY	266	AsnaSPGLTThSerArgHisAlaLeuLeuArgGluThrValGluInserPro	285	JOURNAL	Disease in Plants
		:::		REFERENCE	Unpublished
				AUTHORS	2 (bases 1 to 11330)
Db	140530	GACGCCGAACCGAGTC-GACCTTTTGGATCGCGACGAAGGCTCCACGCCGA	140648	TITLE	Hinkle,G., Slater,S.C. and Goodner,B.
OY	266	LeuYrGlnLysIleGlnThrGlyArg-IleAspTrpGlnSerValGlnThrArgLeuI	305	JOURNAL	Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
		:::			Cambridge, MA 02139, USA
Db	140649	GTTTATGACGACGACAG- - -CGACGTACATTGGGTGCAC- - -CGCGTCCGGTGAT	140699	COMMENT	On Oct 18, 2001 this sequence version replaced gi:15156715.
				FEATURES	
OY	305	eSerAspSPProAlaLysGlyLeuAspArgSPArgLysProPrioIleAlaGlyAr	325	source	1. 11330
Db	140700	CTCTACACCTCCGCGAGAG- - -CTCCGAGCGTCGAGACCGCGGTGGCTATGAGGA	140756	gene	/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
OY	325	gluGlnAspAlaLeuLysGlnProGluLysSerValYrLeuValSerProTyRheVa	345		/strain="C58"
					/db_xref="taxon:181661"
Db	140757	GTTGCTCCCAATCATTCAGTCCGCCGACGCGCAAGCGCAATCGTCCCGTATTTAT	140816	CDS	331. 1221
					/gene="AGR_C_3001"
OY	345	lProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValTh	365		331. 1221
					/gene="AGR_C_3001"
Db	140817	CCCCGCAAGAAAGGTTCAAAATCTCGCGACCTCGTCGACGAGAGTCAGATCCG	140876		/note="probable rmlB3 protein - Mycobacterium tuberculosis (strain H3/RV)"
OY	365	rValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyRVally	385		/codon_start=1
					/transl_table=1
Db	140877	TGTTTGAACACATCTCTGGCGCGACGATCGCGCGCTCCAGAGCTACGCCCA	140936		/product="AGR_C_3001p"
OY	385	STYrArgLysProLeuLeuLysAlaGlyIleLysLeuTyRGluleGlnProAsnHisAl	405		/protein_id="AK87405.1"
					/db_xref="GI:15156716"
Db	140937	CTATCGCAAGCCCTGTTGAGAAAGGCGTACAGCTTCGAGCTTCAGCCCTTAGTAG	140996		/translation="MRTVLVTSGRVGRATFSLAGRHDTGIDRSPFTTHTYDPA
OY	405	aValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaL	425		DESTLRAFERADAVIHTALAHAPVCCVPAEERQRLNVGARVAAMAATGVRLV
					FTSTALYGHAVSSGCTFIDEDPPQPSLYHRTKLEAEHLEEMAGPHLAVRLV
Db	140997	GCAGCCCAAGAT- - - - -TCTGTATTGGATCAAGGCGCGACCTCCACACCA	141047		SRSFPEADVAAYAROHGVDIRVDAAHVATLACAGEDFORYVTSSTFFADDCDA
OY	425	sThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSe	445		LAKDAPSVLROTRPGLDAPFAOEQMALPTTIDRVYSPARAVEIGLGMVSGCSDEVLA
					LARSILVLPAGNATIGRSE"
Db	141048	AGCGTTCAGCCGTCGACGACGAGTGGTTCGTGGGATCATTCATTCGATCCGCGTC	141107	gene	complement(1229. 1837)
OY	445	rAlaLrGluAsnThrGluMetGlyValValIleGluSerProLys- - -IleAlaGluCl	464		/gene="AGR_C_3003"
					complement(1229. 1837)
Db	141108	CGTATCGTGAACCTCGAAATGGAGTGTTTTAAAGACCAAAATCGTATGGAAT	141167	CDS	complement(1229. 1837)
OY	464	mMetLysrGlnLeuAlaAspThrSerProGluTyRArgValThrLeuAspAr	484		/gene="AGR_C_3003"
					/note="(AF218420) c1p"
Db	141168	GGGCGACCGCTCAATCCGACAGCTCGCGCGACAGCAAGCTACGGCTGAACCTC- - -AA	141224		/codon_start=1
OY	484	gHisAsnArgLeuGlnTrpHis- - - - -AsPProAlaThrArgLysThrTyRProAsnCl	502	gene	/transl_table=1
					/product="AGR_C_3003p"
Db	141225	GAAACGGGGTTCGCGCTGGCATGCGACGATGAAGGCAACCGCAACCTATGTCGCGA	141284	CDS	/protein_id="AK87407.1"
OY	502	uProGluAlaLysLeuTrpLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGl	522		/db_xref="GI:15156718"
					/translation="MNROIKVPDAIELEYDLDEPPQVMRAISIPAREHMLPKAL
Db	141285	ACCGAGCGCAGCATTTGTCGCCGCACTGTCGCCCTGGTCAAGTACTCCGATCGA	141344		AQPAIKVSPGRCVRLREDTPPELESTYTFQIASNDRCGTSLRIITHDLPALKSGM
OY	522	uSerLeuLeu	525		TKAANHSPIKLAA"
					2704. 4260
Db	141345	ATCCACGCTT	141354	gene	/gene="AGR_C_3007"
					2704. 4260
RESULT 16					/gene="AGR_C_3007"
AE008085					/note="HYPOTHEICAL 55.9 KDA PROTEIN IN CSGC-MDOC
LOCUS					INTERGENIC REGION"
DEFINITION					/codon_start=1
					/transl_table=1
ACCESSION					/product="AGR_C_3007p"
AE008085					/protein_id="AK87408.1"
VERSION					/db_xref="GI:15156719"
KEYWORDS					/translation="MTAFVFIITLILMLIGPSLFVIGKOREKAIRKRPSTALPYED
SOURCE					ETALDRHQSTIRNGWENKALCLHSNLDFAVAVAAARAAGRLDLMYVMWADLNG
ORGANISM					RLMEVETIAADRCVRYRLDLDGVSMSDRIFRAIDSHNIELEFNPTPARNLILH
REFERENCE					KSLEIVRFSVSNRMKNKAMIVAGRAVIVGGRNIGDAPDAERANHFDPDILGFCG
AUTHORS					IVADATILFDYWNMSAVSVPRSLARRPKNKLAKRLREDALDQSEARARYLLEVESQ
TITLE					

```

gene
CDS
    YGRHFLMSDRHLWVDTADVLADPPKAAGRKRRKGNHLMESLLPLMQAGESLHTS
    PYFIKQGVFEFLDLAERGVSALITNSLAATDYAAHAGYARRKPLSGVYLAH
    LRSODGCSFLIRGSGDGLSKATRTDGEYIGSLNPDPSASLNTNEMGVFNSAP
    LVARDELFAEIRRTMSFEIDISANRYMWTERRGPKTIRRPDAISRRIAGI
    MRVLPLESOL"
    complement(4275..5189)
    /gene="AGR_C_3010"
    /complement(4275..5189)
    /gene="AGR_C_3010"
    /note="(U67070) response regulator protein"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_3010p"
    /protein_id="AAK87409.1"
    /db_xref="GI:15156720"
    /translation="MAKVLMLILKNEPDMSSHTIGRGRNFIETGMOYSGVPT
    NNRILLADSNTQOMYSMLNEMIGVSEVCRSEELQACEHSPPEVTLAISINL
    FGAERGALEYLIDISTPTIVFTSTFEATRETLAKVDLIDKNTFAVDMLESY
    CRFLNHRHVALIVDDSEPTARALLSRKRYNFRVLSAKALEILRANDIGLV
    TDYMPDIDGFELETRIRTVRGSHELRIIGVSSSTNRLLSARFLKAGNDMLRPFD
    EEFYCRVNONLDTLVQIOLAKAGIRPA"
    complement(5267..7780)
    /gene="AGR_C_3012"
    /complement(5267..7780)
    /gene="AGR_C_3012"
    /note="(D1METHYLYCINE DEHYDROGENASE, MITOCHONDRIAL
    PRECURSOR (MEZGLYDH)"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_3012p"
    /protein_id="AAK87410.1"
    /db_xref="GI:15156721"
    /translation="MKTHARAVVIGGVGVSTLYHLAKKQMSDVLIERKELTSGST
    WHAGLPLFMMSYVGOIHKYSVKEVEBELQETGMNMGFSKYSITRLARKDRDDEY
    MYAGIAETIGVRVNMILPREYKEIETPCEITDGLGAIQHPDGTIOTADLQALAKG
    ARDKATITRYNTVTAILEQLEGHKVTVDGEIIAEHISTGFAKRTGEMVIDI
    EPIVEHOYIYTERPALIERROGLPEMGVLRSDSAMNMEEGGAILDPEYGANP
    VCYVDPDSEYELFOBELDLRMHETIAWARYVAFEVGIKVVNGAIALTPYEGNP
    IVGPAPGLKNMNLNGHSFGITAAAGAGMOLAENVDEPTLDMGVDPRRGPATE
    GYLIAKNEEAVANVTMTHYDPDEERSAARPLTTPYDRILKKGIVGSGVSGMERNA
    APEGALREBDLGCADYITSHNAPRIIDDRIVKMSFRSNYFEHNGNEKANTNN
    VGVDMSAFAKMEVSGGARAMDLSILNATPKKRGRIALHLTPNGSVAKIEPTVE
    WAPGFYVSAAGLEADHDVRLAPLDVSGSVLQPIQKGVLAIPKPSDLKLKL
    TRTSLKNDPMLTGKQISVATNAHARVNFVGLGELHHPLEMOYIEPRLEAG
    AEEGKIPGIRAMYSMLEKSYRNMGRRELSYVNGSLDELPELPEKSGIRDLVA
    YKGAIGKVSFTIYVSGNTDVARGSEALISDENGVALARVSGSGMIGKSIALML
    KPEYAAVTKIKIRLGLTYDAEYVEESPTEALLLA"
    complement(7945..10590)
    /gene="AGR_C_3014"
    /complement(7945..10590)
    /gene="AGR_C_3014"
    /note="hypothetical protein PA1408 (imported) -
    pseudomonas aeruginosa (strain PAO1)"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_3014p"
    /protein_id="AAK87411.1"
    /db_xref="GI:15156722"
    /translation="MDVRLMFGMMNGRWRGRAVAGRGGAAPILVYFALLSL
    WMPALAOETTPAPGAIOASTVEQTSDEKWTDAFISGQVAGGRDVOVLVL
    KGRADGIAADAAANAKLRLARDQIKTRLDLGAPGQPEASLVEERARLAPR
    GEVNAMAGEVATANNAAQISNNITAVRRAALFAALFRTEVSATGLDASSAFLEL
    TNLNAPFSMTAYVNNYKRLPEFGAVMSIMAILFLVGGVFEFGSRDDBRFRPEPS
    YLRISLVAFSTMGOSLSLPLETVTSAFELNENLVNDIAPILFGAMATGGEFVS
    RLSTYIAPTOPERMLLKSNGATISSAYILMALVNLVDILYETISETLSPILV
    VAKSFATITIGLILVFLRPMIGESDDYDTGQRLPRMLVILLRVGGLLIADCL
    TGYVGLAAGLQIVATGAVLATMYIGLSGAISROGADESGASLGRVLAERFSAP
    ALDQAGLAAGLQIVATGAVLATMYIGLSGAISROGADESGASLGRVLAERFSAP
    LEGVLAFAIGYIITRMOKMLDNNVMAAGVADAVRNSKVGIGYGLAVAAITGVS
    SAGINLSIALVASLAVGIGGICONTYNSVSGILLVERPKFGDMVATGTEGTV
    KRISVATIEIFRQOSITVPNSEFINSVGMWTRKRMRELTVSAVSDPOGV
    DILLEVAQRPVLRNPEPHEFLRFQGSDFEIRPHLADISNLAVMNARIALILH
    RFRREGIAIPFPORNLNTHVEGDANPOMLAILLSEGGKAVVGHGTAAPTSATLP

```

```

gene
CDS
    KDGPEUDEAAPANNAATKTPGAK"
    10589..10990
    /gene="AGR_C_3015"
    10589..10990
    /gene="AGR_C_3015"
    /note="hypothetical protein"
    /codon_start=1
    /transl_table=11
    /product="AGR_C_3015p"
    /protein_id="AAK87412.1"
    /db_xref="GI:15156723"
    /translation="MPVSLRRFLFSACGAENHPOARFSGVCLLETFTTDPDA
    TWEEIMESAGIGMAIIGGAVGLWLAQOFKMSNNGVMMNILLIGVIAIVNALIVF
    GIVLGMGLYGLIAGFVGCACILIAIRRRRG"
    11145..11220
    /product="trna-lys"
    tRNA
    BASE COUNT 2374 a 3457 c 3227 g 2272 t
    ORIGIN
    Alignment Scores:
    Pred. No.: 66-54 Length: 11330
    Score: 821.00 Matches: 211
    Percent Similarity: 54.338 Conservative: 84
    Best Local Similarity: 38.86% Mismatches: 190
    Query Match: 30.26% Indels: 59
    DB: 1 Gaps: 16
    US-10-066-551-4 (1-525) x AE008085 (1-11330)
    QY 15 IIESerLeuMetLysThrArgSerLeuIIESerLeuLeuCysLeuLeuCysSerCys 34
    ::: III :::: III :::: III :::: III
    DB 2708 CAGCATTTACGGCTCTTCATCATCATCTTCTCTG-----TGCATTTGGGCGGCGCTGT 2761
    QY 35 SerserTrrLeuProProLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 54
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 2762 TCGTCGTC-ATCGGCAACAGCGGGAAGAACCCATTCCCAACGCCCTCCACCGCTCTG 2820
    QY 55 ProValArgLeuAspAsnIIELeuGluIIEArgIISThrProHis--ThraSgLYLeu 73
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 2821 CCGGTTACCGAAGCAACGAACGGCGCGGAGCGGCACTGGCAATCGATCGAAGACGGCTGG 2880
    QY 74 SerAsp-----IIErLeuLeuAsnAspProHisGluIuaIaIaIaIaIaIaIaIaIa 90
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 2881 AACGAAAGAAATGCACTGCGCTTCAATTCACACCTCGATGCTTGGCGGCTGTG 2940
    QY 91 AIAleuIIIEGluSerIaGluHisSerLeuAspLeuGluIYrTYrIIIErPraGAsnAsp 110
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 2941 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3000
    QY 111 IIESerGlyArgLeuLeuPheAsnLeuValTYrLeuAIAIaIaIaIaIaIaIaIaIaIa 130
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3001 CTGACGGGCGGCGGCTGATGATCGCGGAGGTCATCGCGCGCGGCTGATCGCGGCGGCTG 3060
    QY 131 ArgLeuLeuLeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAIAleuAsp 150
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3061 CGGCTCTCTCGATGATCTCGGCGGCTCTCGATCTCGACCGGCACTCTTCATGCGCATGCA 3120
    QY 151 SerHisProAsnIIEGluValArgLeuPheAsnProPheAIAleuArglys----- 167
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3121 AGCGACCCCAATVCAACATCCCGCTGTCACACCCGCGGCGGGAACATCTTG 3180
    QY 168 TrpArgAIAleuGlyTYrLeuThrAspPheProArgLeuAsnArgPmeHisAsnLys 187
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3181 CATCGGACGCTCAACTGCTGCGCTTCGAAAGCTCAACGGCGGCGATCGCAACACAC 3240
    QY 188 SerPheThraIaAspAsnArgAlaThrIIELeuGlyGlyArgAsnIIEGlyAspGlyTYr 207
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3241 GCATGATGATGCGATGCGCGCGCGGATGCTGCGCGCGGAAATATTTGCGGCGCTTAC 3300
    QY 208 PheLysValGlyGluAspThrValIIEPheAIAleuAspAspIIEleuAIAIaIaIaIaIa 227
    ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
    DB 3301 TTGACGCGCGCGGCAACGCGGCAAAATTCACGATTTCTCGATTCTCGCGGCGGATC 3360

```

```

Oy 228 ValGlyValSerHisAspPheArgTyrTrpAlaSerHisSerAlaHisAsn--- 246
Db 3361 GTTCCCGACGCCAGCAAAATCTTGCAGCTACTG-----AACAGCCCGCTTCCGCG 3414
Oy 247 -----AlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeu 260
Db 3415 CCAGTACGCTCGCTGCTGCGCGCCGACGCAACAACTTCCCAACTGGCGCGCAACTG 3474
Oy 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuArgTyrArgGluThr 280
Db 3475 GACGCCCTCCCGCAAGCGAA-----GCCGCGAGACCTATCTGAGCGG 3519
Oy 281 ValGluGlnSerProLeuArgGlnLys-----IleGlnThrGlyArgIleAspTrp 297
Db 3520 GTGGAGAGCCAA-----TATGGTGTATGATCTTCCATATGCGGACCGCTGCACTG 3573
Oy 298 GlnSerValGlnThrArgLeuIleSerAspProAlaLysGlyLeuAspArgAspArg 317
Db 3574 -----GTGATACGCGCAATGTGCTTCCGATCCCGCCGAAAGCGCGGAAACCC 3627
Oy 318 ArgLysProPro-----IleAlaGlyArgLeuGlnAspAlaLysGlnProGluLys 335
Db 3628 CCGAAGGCGCATATTTCTTCTGATGAAACCTATTGCGCGTATGACGCGCGGAGAG 3687
Oy 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355
Db 3688 ACCCTGCACATTAATCTTCTTCCCTATTTCAATCCCGGTAACAGGAGGTGGAGATTTTCTC 3747
Oy 356 LysLeuValGlnAspGlyLleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375
Db 3748 GATCTGCGCGACGCGCGCTGCGCGCATCCCTACCAATTCACCTGCGCGCAACCGAT 3807
Oy 376 ValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLysAlaGlyIle 395
Db 3808 GTCCGCGCGCTTACAGCGGTTATGCGGCTTACCGGAAACCCCTCTCCGCGCGGATA 3867
Oy 396 LysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly----- 413
Db 3868 CGTTTGCAATGATTCGT-----TCCGACGCGGATCAGGCGACGCTTC 3909
Oy 414 ---LeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLys 432
Db 3910 ACCCTGCGCGCTTCCGCGACGAGCGCGACCGCTGACACCAAGCGCTTACCCGCGATGCGGAG 3969
Oy 433 ArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMet 452
Db 3970 ACCGCTATATCGCTCACTCAATTGATCCCGCTCGCTCAATACGGAATG 4029
Oy 453 GlyValValIleGluSerProLysIle-----AlaGluGln 464
Db 4030 GCGGTGATTCATTCATTCAGCCCGCGTGTGCGCGCATGAGAGATATTCGCCGAGAGA 4089
Oy 465 MetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArg 484
Db 4090 ATCCCGCCCGCAGCATG-----ACCTTGCAACTGATGATGCAGACG 4128
Oy 485 HisAsnArgLeuGlnIleTrp-----HisAspProAlaThrArgLysThrTyrProAsnGlu 502
Db 4129 GCCAACCTATTCGTGATGACGAGAGAGAGGCGGACGCAAGATCTACCGCGCGAGA 4188
Oy 503 ProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGlu 522
Db 4189 CCGGACGCGGAGATCAGCGCGGATCATCGCGCATCATCGCTGCTGCGCGCTTGA 4248
Oy 523 SerLeuLeu 525
Db 4249 TCTCAGCTT 4257

```

```

RESULT 17
AE009120 12593 bp DNA linear BCT 20-DEC-2001
LOCUS Agrobacterium tumefaciens strain C58 circular chromosome, section
DEFINITION 146 of 256 of the complete sequence.

```

```

ACCESSION AE009120 AE008688
VERSION AE009120.1 GI:17740055
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Boye, S.R., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutayvin, T., Levy, R., Li, M.,
McLelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.,
and Nester, E.W.
The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
21608550
11743193
2 (bases 1 to 12593)
Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Zhou, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Chen, Y., Boye, S.R., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutayvin, T., Levy, R., Li, M.,
McLelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.,
and Nester, E.W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA
FEATURES
source
Location/Qualifiers
1..12593
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
80..739
/gene="Atu1624"
80..739
/gene="Atu1624"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=1
/product="methyltransferase"
/protein_id="A1A2627.1"
/db_xref="GI:17740057"
/transl_table="MTRAPLRTDRKRKDVMAQONAEHIADMRGSGERMAAYQARLD
VQSAARLFAALFKGSDLLIARFDEARADNEIESYCSRAVAMPMAHAGDDHVCAP
G"
721..1632
/gene="Atu1625"
721..1632
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=1
/product="methyltransferase"
/protein_id="A1A2627.1"
/db_xref="GI:17740057"
/transl_table="MTRAPLRTDRKRKDVMAQONAEHIADMRGSGERMAAYQARLD
VQSAARLFAALFKGSDLLIARFDEARADNEIESYCSRAVAMPMAHAGDDHVCAP
G"

```





QY 74 Serasp-----IleTyrLeuLeuAsnAspProHisGlnAlaPheAlaAlaArgala 90  
 Db 4782 AACGAGAAATGACATGCTGCTGCTTACATCCAACTGAGTCCCTTCCGCTGTG 4841  
 QY 91 AlalaleuIleGluSerAlaGluHisSerLeuAspGluTyrTyrIleThrArgAsnAsp 110  
 Db 4842 GCGGG 4901  
 QY 111 IleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgVal 130  
 Db 4902 CTGACCGGG 4961  
 QY 131 ArgLeuLeuLeuAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAsp 150  
 Db 4962 CGGCTGCTGCTGATGATCTCGGCTCTCCATGTCAGACCGCATCTTCATCCATCGAC 5021  
 QY 151 SerHisProAsnIleGluValAlaArgLeuPheAsnProPheValLeuArglys 167  
 Db 5022 AGCCACCCCAATATGCAACTCGGCTGTTCATCCACCGGGGGGCAAGAAACATCTG 5081  
 QY 168 TrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgMetHisAsnLys 187  
 Db 5082 CATCCACGCTCGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5141  
 QY 188 SerPheThrAlaAspAsnArgAlaThrIleLeuGlyArgAsnIleGlyAspLeuTyr 207  
 Db 5142 GCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5201  
 QY 208 PheLysValGlyLeuAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerVal 227  
 Db 5202 TTCAGCGCGCGGCAAGCGGCAATTTCCACGATTCGACATTCGATTCGATTCGATTCG 5261  
 QY 228 ValGlyGluValSerHisAspPheAsnArgTyrTyrPheAlaSerHisSerAlaHisAsn 246  
 Db 5262 GTTGCCGACGCCACCAATCTTCGACGATTAATG-----AACAGCGCCCTTTCCTCG 5315  
 QY 247 -----AlaThrArgIleLeuArgSerGlyAsnIleGlyLysGlyLeu 260  
 Db 5316 CCAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5375  
 QY 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThr 280  
 Db 5376 GACGCGCTGCGGCAAGCGAA-----GCCGAGACGACCTATCTGAGCGG 5420  
 QY 281 ValGluGlnSerProLeuTyrGlnLys-----IleGlnThrGlyArgIleAspTrp 297  
 Db 5421 GTGAGAGCCAA-----TATGCTGATGATCATTTCTGATGCTGACCGCTGCTGCTGCTG 5474  
 QY 298 GlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArg 317  
 Db 5475 -----GTGATACGCGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5528  
 QY 318 ArgLysProPro-----IleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLys 335  
 Db 5529 CGCAAGGCGCATTAATTCCTATGTAAGACCTATGCGCGCTGATGTCAGCGGGGGAAG 5588  
 QY 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355  
 Db 5589 AGCTTCGACATTAATTCCTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5648  
 QY 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375  
 Db 5649 GATTCGCGGCAAGCGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5708  
 QY 376 ValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLysAlaGlyTyr 395  
 Db 5709 GTCGCGCGGCTTACGCGGCTTATGCGGTAACCGGAACCGCTTCTCTCGCGGGGTA 5768  
 QY 396 LysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly----- 413  
 Db 5769 CGTTTCATGAATTCG-----TCGACGCGGATCAGCGACGCTTC 5810

QY 414 ---LeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLys 432  
 Db 5811 ACCCTGCGCGGTTCGGGACAGCGCAAGCCCTGCACACCAAGCGCTTACCGCGGATGGCG 5870  
 QY 433 ArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMet 452  
 Db 5871 ACCGCTATATGCGCTCACTCAATTCGATCCCGCTCGCTGCTCAATTCGGAATG 5930  
 QY 453 GlyValValIleGluSerProLysIle-----AlaGluGln 464  
 Db 5931 GCGGTGCTTCAATTCAGCCCGCGTGGGCGCCGATGAGACAGATATTCGCCAAGAA 5990  
 QY 465 MetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArg 484  
 Db 5991 ATCCCGCCGACGATG-----ACCTTCGAATCTGATTCGACAGC 6029  
 QY 485 HisAsnArgLeuGlnTyr-----HisAspProAlaThrArgLysThrTyrProAsnGlu 502  
 Db 6030 GCCAACCGTATGCTGTGATGACGAGACAGCGGGGCGAGCCAAAGATCTACCGCGCGAA 6089  
 QY 503 ProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGlu 522  
 Db 6090 CCCGACGCGGCGATCAGCGCGGATCATCGCGGATCATGCTGCTGCTGCTGCTGCTGCTGCTG 6149  
 QY 523 SerLeuLeu 525  
 Db 6150 TCTCAGCTT 6158  
 RESULT 18  
 AP003013 347950 bp DNA linear BCT 15-MAY-2001  
 LOCUS Mesorhizobium loti DNA, complete genome, section 20/21.  
 DEFINITION AP003013 BA000012  
 ACCESSION AP003013.2 GI:14027324  
 VERSION  
 KEYWORDS  
 SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.  
 ORGANISM Mesorhizobium loti  
 Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 Phyllobacteriaceae; Mesorhizobium.  
 REFERENCE  
 1 (sites)  
 Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,  
 Watanabe,A., Idegawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,  
 Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,  
 Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,  
 Takeuchi,C., Yamada,M. and Tadota,S.  
 Complete genome structure of the nitrogen-fixing symbiotic  
 bacterium Mesorhizobium loti  
 JOURNAL DNA Res. 7 (6), 331-338 (2000)  
 MEDLINE 21082930  
 REFERENCE 2 (bases 1 to 347950)  
 AUTHORS Kaneko,T.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research, Yana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:kaneokkazusa.or.jp,  
 URL:http://www.kazusa.or.jp/rhizobase/  
 Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
 On May 11, 2001 this sequence version replaced gi:11994988.  
 COMMENT  
 FEATURES  
 location/Qualifiers  
 1..347950  
 /organism="Mesorhizobium loti"  
 /strain="MAFF303099"  
 /db\_xref="taxon:381"  
 /complement(78..1037)  
 /gene="m117923"  
 /complement(78..1037)  
 /gene="m117923"  
 /codon\_start=1  
 /transl\_table=1  
 /product="transcriptional regulator"  
 /protein\_id="BA53594.1"  
 /db\_xref="GI:14027325"

```

/translacion="MRLISQVNLNLSLKIVESAARHGNFTFRAGEQFITASAVQRFKS
LEDOLEKFIPOFGGNAVSLTEGETYVSRVREALERIYAASMEATGSOEHLKISVL
PTPAARMLFPRLEPFOFRYPDIIVMRVSTSYATHESTSDPDLETIRYGDHNGILPSDL
LFREDLITPVCSRKLFHEYLGDKPVSKVPPDLRHFTLLHSDTCTONQSWIGFAGASF
VLSEKTSIYFDSCMSYEAANAGMGFAANRAYSADIRAEKLVAPFAVHHNPSAGWY
FVSPKSLCAARKVLLFKQWMAEALRTOROLDSELRDAEAV"
/gene="mlr7924"
1150. 1986
/gene="mlr7924"
1150. 1986
/gene="unknown protein"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53595.1"
/db_xref="GI:14027326"
/translacion="MNAHTIPELRIAMSREALIGHETAKVSSQEVAYLHGYDALL
AIEEAALATKDSHAHVHKHDLTFTTGADRTPEIKELHLKRLIERLSDMGKTLER
YPLSTVSTVTFDSHAFNRDGAEMHCDGVPTELPLISNPLVGHLLEYCDSDSETGRA
ILSEGREIPRRIRIMRIDHKMNYATLGQFLGVLHRTAPLOYGERVTLLEYNORSVAKPY
DONRMFYLAADNDHREXVNELAEDVMTNQLPAYRREAEHPTPAPASETOVSGARG
SW"
2032. 2958
/gene="mlr7925"
2032. 2958
/gene="mlr7925"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53596.1"
/db_xref="GI:14027327"
/translacion="WQHRDRAPGRLPFRHSSSISALVFAAGAVALWMTNALVGRKSL
ATHPSVOFLQFAGALVFPALIRMSREAPSPVAPRAVALAIGIIGLVGMLOY
IAPFSPYTEANLVATWPLVAAVIAFNDRRRALGLAIVGFVGLVLEIYVSGRE
NNMRGDLVGYFAAGSALCMAPFSVMGRATSPDRLLLSIYGVALLALIMSGRG
VAMRAGMDLALGLYLGAGPKLIGYFMSRAKLSESGKAVAYVAILTPASTILLTLSG
EQLTTAAGAVLIGSCTAVGLERSEADYV"
2951. 3154
/gene="msr7926"
2951. 3154
/gene="msr7926"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53597.1"
/db_xref="GI:14027328"
/translacion="WEDDNERLARQEAHMLIKEFGAEPVYAAMKAKEAIQKXGRC
ARMKRIELADARTTKSAASKY"
3307. 3759
/gene="mlr7927"
3307. 3759
/gene="mlr7927"
/note="unknown protein"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53598.1"
/db_xref="GI:14027329"
/translacion="MKELSNIEDKTVIELTADIVASAYVGNPPLPASGLPDLIASVAS
VRKLAGAVVESPSLVPAVNPKEKSVFDPYIICLEDGKKFKSLKRLLRDYGSLSPDIYR
AKWGLPPDYPMWAPNYSATRSKLSLSTGLGRPAAPAAVAKKAKAKA"
3916. 4065
/gene="msr7929"
3916. 4065
/gene="msr7929"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53599.1"
/db_xref="GI:14027330"
/translacion="MNCNAYFEAKGRKSALVLDKMTLLELLGSGKPLDRNQCRA
SLRPR"
complement(4476. 4730)
/gene="msl7931"
complement(4476. 4730)
CDS

```

```

/gene="msl7931"
/note="unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB53600.1"
/db_xref="GI:14027331"
/translacion="MRSERSLIMISMAALLHRTSTLRKALRASWPDRVKTGSGSALRH
OPGKIKMRSDPCOVADRKTATGTDIEDTGLRNLYYFVSC"
4699. 6576
/gene="mlr7932"
4699. 6576
/gene="mlr7932"
/note="ABC transporter, ATP-binding component"
/product="ABC transporter, ATP-binding component"
/protein_id="BAB53601.1"
/db_xref="GI:14027332"
/translacion="MLIINDLSLRNAGRLLDHASLTLPAGTACLVGRNGTKTTLF
KAITDPFSEISLSPKNTIRIGVYAOPAGEEPLIEYLKADVERTALLKEETAT
PBRITADIHMLRADIDAHSAESRAATILAGLGDPAAGSPASSFGGMRMVALAAV
LESPDILLIDEPYVLDLEGTIMLVNYSKYPHTVLLISHROILNRAVSYVALDO
KLTTFWRGCIYDOFEQYTBOKETLDEKGVKQDPAKIHMESEYERFRASAKARQOSR
IKALEMKRPIAIVNDVSRPESEPEPVKTVASPIVALNNVNGITEGPILKMTLRI
DADRIALLGANGNGSTFAKLISGRLOESQMTVAEGLKVAIFAQHQDLDRPEN
AYEHRILMPEAPESEKVRGVAQFGLTTEKMTPAKDLSGEKAFLMGLSAFEGPNL
FLDEPTNHLIDISRESLIHALNEPQGVAILISHRHILEATADRLMLVKQCAVNPD
GDLEDKYKTLVGVSGDRGKREADKASADRDRDAARAAFEPLAKETIRATEALMDR
IRKRIDGIEDELANAIYEKDPSTATRLAKERSQLAQTLAGHEKMLMSAEYEGTA
E"
6667. 8178
/gene="mlr7933"
6667. 8178
/gene="mlr7933"
/note="similar to formate dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB53602.1"
/db_xref="GI:14027333"
/translacion="MNOHAKLRIGHSACPHDCPSTCALEVELLDGNRIGRVAKANT
YTSGVCAVARYADRVHHPDRLLKPLIRAGAKKGGVNOEASMEALDLVAEKFTAE
EKYSEYVWPYRYFASCTAAYVGSQYGGGAFHNSGIFKINCDELCTGRDRDSIR
YLDHSRIGVLTGASDALTGGPPVTAMLIONTNPNVAPBEQVLVQGLRDLFACVH
EQEMTDTRKLAADVLPATMFEHEDVDVYGGGQHTLTPKRLLEPEGGPTNFIYEQ
AERLGVADRPGGLVQOQHIDIILKRGIGSGDSLEKQWDLQDPFTAHFIKGFH
SDGKFRFRADMTGQAAPNRPKSMGLPEVEVLPEPDPHVDIEVAADAHPRLLTSP
ARNFLNSTPAETPVSKKEKGRPALLLHPDDADGLADGDRVEVGNORGDDVLHAKFF
DQIKRGVYIABGIMPNSAHREBEGINVLGADAPAPYCGAAPHDKVMLRKAG"
complement(8203. 9390)
/gene="mlr7934"
complement(8203. 9390)
/gene="mlr7934"
/note="similar to dimethylaniline monooxygenase"
/codon_start=1
/transl_table=11
/protein_id="BAB53603.1"
/db_xref="GI:14027334"
/translacion="MELTFSGRELSREARLDOTTVAIIGAGPACILAAVACLRQAGVD
FLIIRKEOQAAAMBRHYERVHLHTTKRSSLPVPPKHYRYPVRAALFVYLDAYA
QRFDLRPOGETVYKAVTODGRKMRDASGPIRADVYIASYNAEPLRAPAGIDTF
TGKTLHSADYRNAKFPAGOSVLVIGMNGTGAETIALDLLENAGQPTISVGGVHAIPIRE
LGVPDIOWGMAARLPORLNDALRPVLLIDLVLGKDRKGLRPOGLLEQTAIVASRI
PVIDGTTGIREGAIKVAIPDITTEISORARFADKNGHGFDAIIFATGVRPQYARFLE
PGIOWDRSGVTPKASDLGLYLIGFNNAVYGLLRIGIEQAQIADDIRHLNRKKAIEI
LPV"
complement(9426. 9905)
/gene="mlr7935"
complement(9426. 9905)
/gene="mlr7935"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53604.1"
/db_xref="GI:14027335"

```















Db 208382 GATTCTTCGATCGTCACCGATGTCGCTGGAGCGCGAAGCCGACCTCCACACC 208441  
 Oy 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444  
 Db 208442 AAGGCTTCGTCGTGATGCGCTCGCGGTACGTCGCGCTCATGATTCATCCGCT 208501  
 Oy 445 SerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluGln 464  
 Db 208502 TCCTGGCTGCTCAACACCGAGTGGCATGCTTTCCTCCAGACGAGCGCTCCGCCGAG 208561  
 Oy 465 MetGluArgThrLeu---AlaAspThrSerProGluTyrAlaTyrArgValThrLeuAsp 483  
 Db 208562 CTCGATGATATTCGACGAGGAAACGGCGCGGACGAGCTATGCTGATGCTG--- 208618  
 Oy 484 ArgHisAsnArgLeuGlnThrPheAsp-----ProAlaThrArgLysThrTyrProAsn 501  
 Db 208619 AAGGACGGCGCGCGCGATGGGGATCGGACGACGAGCGCTCGCGTGGCGGCGC 208678  
 Oy 502 GluProGluAlaLysLeuThrLysArgIleAlaAlaLysIleLeuSerLeuLeuProIle 521  
 Db 208679 GAGCGGACGAGCGGATCTCCGCGGCTGATGCTGGCGTATCCGCTTTCGCGCTTG 208738  
 Oy 522 GluSerLeuLeu 525  
 Db 208739 GAGTCACAGCTT 208750  
 RESULT 22  
 AE001456/c 12892 bp DNA linear BCF 20-JAN-1999  
 LOCUS Helicobacter pylori, strain j99 section 17 of 132 of the complete  
 DEFINITION genome  
 ACCESSION AE001456 AE001439  
 VERSION AE001456.1 GI:4154689  
 KEYWORDS  
 SOURCE Helicobacter pylori j99.  
 ORGANISM Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 Helicobacter.  
 1 (bases 1 to 12892)  
 Alm,R.A., Ling,L.S., Molr,D.T., King,B.L., Brown,E.D., Doig,P.C.,  
 Smith,D.R., Noonan,B., Guild,B.C., deJonghe,B.L., Carmel,G.,  
 Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,  
 Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.E.,  
 Voyis,G.F. and Trust,T.J.  
 Genomic-sequence comparison of two unrelated isolates of the human  
 gastric pathogen Helicobacter pylori  
 Nature 397 (6715), 176-180 (1999)  
 9923682  
 2 (bases 1 to 12892)  
 King,B.L., Alm,R.A. and Trust,T.J.  
 Direct Submission  
 Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney  
 Street, Cambridge, MA 02139, USA  
 Address all correspondence to: hpearb.us@astracom or Richard  
 A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,  
 MA, 02139. Lo-See L. Ling, Donald T. Molr, Douglas R. Smith,  
 Brydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,  
 Rene Gibson, and Gerald F. Voyis are with Genome Therapeutics  
 Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and  
 Diane E. Taylor are with the University of Alberta Department of  
 Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G  
 2H7 and the Canadian Bacterial Diseases Network. All other authors  
 are with Astra Research Center Boston, 128 Sidney Street,  
 Cambridge, MA, 02139. Putative identifications, sequence  
 alignments, and name and sequence search capability are available  
 at ARB's World Wide Web site. (URL:  
 http://www.astracom.com/hpylori).  
 Location/Qualifiers  
 1..12892  
 /organism="Helicobacter pylori j99"  
 /strain="j99"

gene  
 CDS  
 /db\_xref="taxon:85963"  
 complement(227..1930)  
 /gene="jhp0174"  
 complement(227..1930)  
 /note="jhp0174"  
 /note="similar to H. pylori 26695 gene HP0186, HP0187, and  
 HP0188"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative"  
 /protein\_id="AAD05767.1"  
 /db\_xref="GI:4154702"  
 /translation="MKTTKITEFOEGSTIPNORDYAMKDKNFRIDMEDLEAIEYN  
 KYGGEHPIGVWVAKNEDNKKLYDIDGQQFTTITMILHLVLSKQYEEKRRETRKL  
 KQGEHLLEVASQNSEFFKTLLEVAEKGNSHCEKADDEGKMLFEVLKALIDISK  
 LSEEEVNERLEVLKMYLMLEEDPKAITFQSVNDRGVPLLDLRLKLSLIYSN  
 TFGDKRGIDQFIIDHGEIKRIFAKIKKSHISSGAFENEGDIRYVAGSGRFG  
 IEFGLHYRTSDNTPYKTKAEIKTKSKLSFQSVSDLNKRYOAFILDSIDIN  
 PTTKVMILNPLFNSLRTRKINNELDETELAKRAKTDIVCFKSTRYQSAAYL  
 INAYLRKGESEKSGMTAQCRNDLELAFROSVNAEFSSCFHYEFERKQCEGLADL  
 KLLPGKQFSQEKELIPIINLDIDNEIEIQGLGEDRKDLRAVITYGNLSLEKPL  
 NIKGKDKDLVEKDEIKYSESEIPFRNRNVAQFNKKVILTRNDEQEWLIDTFKDFAT  
 H"  
 2086..2619  
 /gene="jhp0175"  
 2086..2619  
 /gene="jhp0175"  
 /note="similar to H. pylori 26695 gene HP0189"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative"  
 /protein\_id="AAD05759.1"  
 /db\_xref="GI:4154694"  
 /translation="MLEKLETERVLPATRWMLAPCIAMSLVLYGVFKMELMHMS  
 HLDITSEDLVLSAGVDLDFMAGLVLMVLAESFEVSKLDVDSSETWMLKHTGF  
 NALKKQSTIVASIAIFLKRKMSLEDVLSIRKPIPLSHNPITFQGVYHLVVCSA  
 LITAVTNNAFSQKERH"  
 complement(2626..4134)  
 /gene="jhp0176"  
 complement(2626..4134)  
 /gene="jhp0176"  
 /note="similar to H. pylori 26695 gene HP0190"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative cardiolipin synthase"  
 /protein\_id="AAD05760.1"  
 /db\_xref="GI:4154695"  
 /translation="MKIEFVLISVFFNCGFLAYKTPISNPISPYPTTIGSLVA  
 NKIKENPKHSAAILLEDGPDALHRLVGLIRMSOKSIDMORTIYKNDLSQVIKELN  
 AANGVAVRIILDNGDSDPSDITMLNPFKNITVYKFNPIYITNKKLRIFEMADYE  
 RIKRHMHKLEFVDNFAVILIGRNIGDVFENDIDTNFLDIDLFPQGVASKASEFE  
 NYWRHRSIPVSLRTHRLKNNVKEIAKLEKIPISAEDANEFEKRVNDFIERFOY  
 QYPIYGNATIELADLPKIDTLPYSPIKIAFEKALKNAKDSVFASSYFPGKIMKI  
 FKNQISGIELNITLNSISSTDAIVVYGAMRYRNNKLVGMQANVEIIRNDFENQIG  
 RFTSKSLHGKTIYFDALITLGSFNDIPRSAYINTEFSAYFNDSPAKRRRLSLKH  
 AQSWMHLVLRHRYIWEATEBERGILIHENKNSDTSFFLLIKEMSKVLPKRL"  
 complement(4156..4893)  
 /gene="frdb"  
 /note="jhp0177"  
 complement(4156..4893)  
 /gene="frdb"  
 /note="similar to H. pylori 26695 gene HP0191"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Fumarate reductase"  
 /protein\_id="AAD05761.1"  
 /db\_xref="GI:4154696"  
 /translation="MSDNERTIVRVLKFPDQSAVKKPHREKIQAKTSPMTFLIALN  
 LIREHDPDLSDFVCRAGICGSCAMNVEVNPRIKACTLTSSENGVITLMPMEGGL  
 IKDLSVTQDMFSDMTKRVSMAKSEVNIITREKREVDDEQVEYELREKICPSCFC  
 IASCGTKLMPNFIAGAAMNRMFRKSSIDERSDDDFEYLVGDDGDFGCMSLIACH  
 DTCPKELQLOSSIATLNRMLKVGKSK"  
 complement(4886..7030)

```

CDS
/gene="frdA"
/Note="fhp0178"
/complement(4886..7030)
/gene="frdA"
/Note="similar to H. pylori 26695 gene HP0192"
/codon_start=1
/transl_table=1
/product="Fumarate reductase"
/protein_id="AAD05762.1"
/db_xref="GI:4154697"
/translation="MKITTCALIIIGGGLASIRASICKQKGLNTIVLSIVPRRSH
AAOQWQASLANAKSSEGDNEHLFDITVGSQWQOVARMEVTTAPKAIKELAS
MGVPMTRIKKGDRAVNVGEHVIITRDGTYILSRDGGTKKWKCTFTADGCTH
LYAANKEALHHRVDIQRKDLAFIHDNKCQAVRDLIGELISAYSKGLTATGG
YGRVYKHTTNAVICDAGASALETGVKLCNMEAVOPALVSPALVSGITGGCGDG
GVLDRKGRFRMPAYEPERKELASRDVSRRLIHLIOKCYGAKSPGQHWMLDIIIG
RNVYKRLRDVDMPTAMTFAGIDPADSEQGTQDNNOGATNPEVQAAKAKGPIK
PMOYHSMGCVTRNKGETHLKGLFCAGACMDLHGEKEDVYEIREKMEVNDKGVF
FASHCLEQLEINTQKVEAFIKESODYMHFLHNEGKDYVEIREKMEVNDKGVF
REGKRLLEALQELQELVARSKNICVKNVLLHNELEDEYATKMKLALCITOGALL
RTESGATRIDYQKRDDEKLNRTLASMPSEADMDPTIEYEDVMKMEISPDGKY
GKSNFIHPKKEERDAEILKTIIELEKIGDRILEVQALMPFELQEKYKARNRLED
EEVARGEHLYSFVHDLIDQHNANKGEHNE"
/complement(7040..7807)
/gene="frdC"
/Note="fhp0179"
/complement(7040..7807)
/gene="frdC"
/Note="similar to H. pylori 26695 gene HP0193"
/codon_start=1
/transl_table=1
/product="Fumarate reductase"
/protein_id="AAD05763.1"
/db_xref="GI:4154698"
/translation="MOOEELIEGYGASKGLKSGIYAKLDFIOSNGTLALFMIAH
MEIYSLISDEAMTKVAKFPEGSLFLKAGSPAYSVAAVITLIVAHARLAKRP
INIKQYKFKTHLKHMGHDSLMFTQALGTFAPFSLISLFWLEPEISIGRSS
YRVTQNMWLLYIFLFAVELHSGISGLRIALIKMGWFNVSIQGLRKIKMAVSFEIV
GLICTGYAVIKKGLKNGIKTMOEALIEADGKPHKE"
/8017..8721
/gene="tpi"
/Note="fhp0180"
/8017..8721
/gene="tpi"
/Note="similar to H. pylori 26695 gene HP0194"
/codon_start=1
/transl_table=1
/product="TRIOSE PHOSPHATE ISOMERASE"
/protein_id="AAD05764.1"
/db_xref="GI:4154699"
/translation="MTKIAMANFKSAMPEFKSHAULKELEKTLKPOHDRVFPDFL
GLPNALFHTFLGQNAVPRDCAFTGITSKHLELKHILLGHSRRVLKESPS
FLKRPDEFKGNKRIYCIIGEDLTTREKGRVAKPELSEDLNIDNYSMLIYAV
IMATGTSKASLEDIYLTHGFLKQILNKPTLLYIGGSVNAONAKELIISDVGLLIG
SASLELEKRTITISFL"
/8731..9538
/gene="fabI"
/Note="fhp0181"
/8731..9538
/gene="fabI"
/Note="similar to H. pylori 26695 gene HP0195"
/codon_start=1
/transl_table=1
/product="ENOYL-ACYL CARRIER PROTEIN REDUCTASE"
/protein_id="AAD05765.1"
/db_xref="GI:4154700"
/translation="MGFLKKGKGLIVCAVANKKSTAYGAGSCFNOGATLATYINLES
EKRVPIAOELNSPVYELDYSKEHFSKSLNNIKODGSLDFIVHSAFAPKALLES
SLLETSKAFYTAAMEISYSLIELNTLKPLNNGASVLTLSYGSTKYAHYVWML
AKALESAVRVLAVDLGKHNRVNAISGPIRTLASGSIADPFMIKLKNEINADLRKN
VSLBEVGNAGMYLLSLNSGVSEVHFVADAGYHVMGMAVEKDNKATLLMDLKEKO"
/9568..10578
/gene="lpxD"
/gene="lpxD"

```

```

CDS
/Note="fhp0182"
/9568..10578
/gene="lpxD"
/Note="similar to H. pylori 26695 gene HP0196"
/codon_start=1
/transl_table=1
/product="UDP-3-O-[3-hydroxymyristoyl]"
/protein_id="AAD05766.1"
/db_xref="GI:4154701"
/translation="MKISELLSANSIETFSNDEPVHAALELDKATPNDISYDQVCY

Alignment Scores:
Pred. No.: 4,31e-39 Length: 12892
Score: 632.00 Matches: 163
Percent Similarity: 51.92% Conservative: 92
Best Local Similarity: 32.93% Mismatches: 184
Query Match: 23.30% Indels: 56
Gaps: 17

US-10-066-551-4 (1-525) x AE001456 (1-12892)
QY 64 IleaYGHISthPrOHISthrasnGlyLeuSerAPileTYr----- 77
Db 4047 ATCTCTTATGACCCCTACACTACACCATTTGGAGCTGTATGCTAAATAATTAAGAA 3988
QY 78 -----LeuLeuAAsnAPPrOHISGLuAlaPheAlaIaArg 89
Db 3987 AACCTTAACATAGTGGCGCCATTCCTTTTGAAGAAGCGCTTGATGCTTGGACAGAA 3928
QY 90 AlAlAlaLeuIIGluSerAlaGluHisSerLeuAspLeuGluTYrIleTPrArgAsn 109
Db 3927 GTGGCTTATTAAGATGATGACCAAAAGCATTTGACATCCAACTTATATCAATAAAC 3868
QY 110 AsPIleSerGlyArgLeuLeuPheAsnLeuValTYrLeuAlaAlaGluArgValArg 129
Db 3867 GACCTTTCTTCAGATGATGCTTAAGACCTTTAAACGGCGCAATGCTGGGTAAAA 3808
QY 130 ValArgLeuLeuLeuAspAsnAsnThrArgGlyLeuAsp-----AspLeu 145
Db 3807 GTGGCATCTCTTTAGATGACAAAC-----GGTTNAGATTCAATTTTTCACACATC 3757
QY 146 LeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPheValLeu 165
Db 3756 ATGCTT---TTAAATTTCCACAAAACATTTGAAGAAATCTTAACCCCTACATATC 3700
QY 166 ArgGlySTPrArgAlaLeuGlyTYr-----LeuThrAspPheProArgLeuAsnArg 182
Db 3699 CGCAAT---AAGGCTTGCGTTATTTTGAATGCTTGGGATTTAGAACGATTAATAAA 3643
QY 183 ArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyIleArgAsn 202
Db 3642 CCCATGACCAACACACCTTTCATCGTGATATTTCGCGTATATATAGGGGGCGCAAC 3583
QY 203 IlegIAspGluTYrPheLysValIleGlyLysAspThrValPheAlaAspLeuAspIleLeu 222
Db 3582 ATTTGGGACAAATTTATTTGATTAACGATTTAGACAGAAATTTTGTGATTTAACCCTTGG 3523
QY 223 AlaThrGlySerValValGlyIleValSerHisAspPheAspArgTYrTrpAlaSerHis 242
Db 3522 TTTTGGGGGGGCTTCAAAAGCCAAAGAACCTTTGAATAATTTATTTGGGATTCAC 3463
QY 243 SerAlaHisAsnAlaThrArgIleIleArgSer-----Gly 254
Db 3462 CCCTCATCCCTGTTCA---TTATTAAGAACCCATAAAGACTCAAAACACGTTAA 3406
QY 255 AsnIlegIlyGlyLeuGluAlaLeuGlyTYrAsnAspGluThrSerArgHisAlaLeu 274
Db 3405 GAAATCGCTAATCCATGAAAAAATCCCTATACCGGCTGACGCAACCAATTTGAA 3346
QY 275 LeuArgTYrArgGluThrValGluGluInsPheProLeuTYrGln---LysIlegIthrGly 293
Db 3345 AAAAAGTCATGATTTTATAGAACGCTTCCAAATAATACCAATACCCTATTATATGCG 3286

```

QY 294 ArgIIeaSPtRgISeSVaI6InthraGleuIIeSeSaSPProIaIaLysgIleu 313  
 Db 3285 -----AATGCATTTTTCAGCCGATTTGCCGCCAAA---ATT 3250  
 QY 314 AspaTgaSPaTgaTgLySProBProIIeaIaGlytArgLeuGInaSPaIaLeuLysGInPro 333  
 Db 3249 GACACGCCCTTGTAATCG---CCCATTCAAAATCGCTTTGAAAAGCCCTTAAAAACGCT 3193  
 QY 334 GluYSerSeVaItyLeuValSerProTyRPhaEaIProthLySserGlythraSPaIa 353  
 Db 3192 AAAGACTCCGTTTTATTCGCTTACATCGTATTTATTTCCAGGTAAGAAAAGTATGAAATC 3133  
 QY 354 LeuAlaLySLeuValGInaSPaItyIleasPaItyhValLeuThraSPaIaSerLeuGInaIa 373  
 Db 3132 TTTAAATAATCAATTTCTAAGGGAGATGAAATGAAATATCTTACCAATTCCTTCATCT 3073  
 QY 374 ThrAsPaVaIaIaIaValaHISerGlytyValLyStyTArgLySProLeuLeuLysaIa 393  
 Db 3072 ACGGACGCTATCGTGTCTATGCGCGTGGGAGGAAAGGTATTCGCAATTAATTAAGCGAATG 3013  
 QY 394 GlyIleLySLeuTyRGLeuGInaSPaIInProaSnHISaIaIaIaProaIaThraLySPaIySgI 413  
 Db 3012 GCGCGGATGCTATGAATAACGAACGATTTTTCACACGCCAGATTAAGGCCG--- 2956  
 QY 414 LeuThrgLySerSeVaItyhSerLeuHISaIaLySthPhaIleVaIaSPaIySArg 433  
 Db 2955 -----TTTAGACCAACCAATTCCTTACACGCGCAACACATTTGTTTGGCATGCCCTA 2902  
 QY 434 IlePhaIleGlySerPhaSnLeuSPaIProaGSeSaIaIaArgLeuSnHISaIaGlyMetGly 453  
 Db 2901 ACGCTTCAAGGAGATTTATATTTATTCGCGCTGCTGCATATATTAACCTAAGAGTGGG 2842  
 QY 454 ValValIleGlySerProLySIIeaIaGInaGInaMetGlyArgTyhLeuAlaSPthSer 473  
 Db 2841 GCTCTGTTTGACACACCCGCTCTTTCGCAAAAGGAGCGCTTTCCTTAAACATATCATGCC 2782  
 QY 474 ProGlyTyRAlaTyRArgValaThraLeuSPaIaGHSaSnArgLeuGInaThraSPaIaSPro 493  
 Db 2781 ---CAACATCATGCGATTTACTCTTGATCGCAT---AGAGATGTTGGGAA----- 2734  
 QY 494 AlaThraTyRLySthTyR-----ProaSnGlyProGlyAlaIaLySLeuThryLys 509  
 Db 2733 GCTACAGAAGAAGCATCTTAATCATGAAAAAAACTCGCTGATACCTCTTCTTTTG 2674  
 QY 510 ArgIIeaIaIaLySIIeLeuSerLeuLeuProIIeGInaSerLeu 524  
 Db 2673 CGTTGATTAAGAAATGCTTAACCTCCTTCTGAAAGAGAACTT 2629

RESULT 23  
 AE000539/c  
 LOCUS  
 DEFINITION Helicobacter pylori 26695 section 17 of 134 of the complete genome.  
 ACCESSION AE000539 AE000531  
 VERSION AE000539.1 GI:2313275  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Helicobacter pylori 26695.  
 Helicobacter pylori 26695  
 Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 Helicobacter.  
 1 (bases 1 to 10631)  
 Tomb/J.-F., White/O., Kerlavage/A.R., Clayton/R.A., Sutton/G.G.,  
 Fleischmann/R.D., Ketchum/K.A., Klenk/H.P., Gill/S.,  
 Dougherty/B.A., Nelson/K., Quackenbush/J., Zhou/L., Kirkness/E.F.,  
 Peterson/S., Loftus/B., Richardson/D., Dodson/R., Khalak/H.G.,  
 Glodek/A., McKenney/K., Fitzgerald/L.M., Lee/N., Adams/M.D.,  
 Hickey/E.K., Berg/D.E., Gocayne/J.D., Uitterback/T.R.,  
 Peterson/J.C., Kelley/J.M., Karp/P.D., Smith/H.O., Fraser,C.M. and  
 Venter,J.C.  
 The complete genome sequence of the gastric pathogen Helicobacter  
 pylori  
 Nature 388 (6642), 539-547 (1997)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 9252185

REFERENCE AUTHORS	TITLE	REFERENCE AUTHORS JOURNAL	FEATURES source
2 (bases 1 to 10631) Tomb J.-F., White O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klein, H.P., Gill, S., Dougherty, B.A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E.F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khaliq, H.G., Glodek, A., McKenney, K., Fitzgerald, L.M., Lee, N., Adams, M.D., Hickey, E.K., Berg, D.E., Gocayne, J.D., Utterback, T.R., Peterson, J.D., Kelley, J.M., Cotton, M.D., Mdelman, J.M., Fujii, C., Bowman, C., Wathey, L., Wallin, E., Hayes, W.S., Borodovsky, M., Karp, P.D., Smith, H.O., Fraser, C.M. and Venter, J.C.	Direct Submission Submitted (06-AUG-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA 3 (bases 1 to 10631) White, O. Direct Submission Submitted (17-MAR-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA Location/Qualifiers 1. 10631		/organism="Helicobacter pylori 26695" /strain="26695" /db_xref="taxon:85962" /complement(228. 1442) /gene="HP0186" /complement(228. 1442) /gene="HP0186" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=1 /product="H. pylori predicted coding region HP0186" /protein_id="AAd07264.1" /db_xref="GI:2313284" /translation="MSEEVNERLEVLWLMFEEPPDGRAIRTPQSVNDRGVPILL LURKSKLLIYVSNIFGCGKRGDQFTIHGGELEKFEAKIKSDHISVVGPEGDI FRHAGSGKDFGEEFGHVAESGDTQYETKIKDELKIKKSKISFQSYSDLNKPYO AFPLDSEIDNPTTFKVMILINKIDDSFNSLITKIKNNELDEDTLLFAKTDVLEK ATDREGTDMLINAYLKKGEGKSEHIAOCRNDIGAFQSVNANSSSCFHTIFE KNOEGMLADLKLIPKROSOEKHEIPIINLKGSENNIRIDGFECDKDLDEYIDT YGNLISLEKSLNRKASDKDLVGKDEIYKSEIPEPRFDTKNEKKALVKNREBREM LIDTFKDFPAAH" /complement(1469. 1756) /gene="HP0187" /complement(1469. 1756) /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=1 /product="H. pylori predicted coding region HP0187" /protein_id="AAd07265.1" /db_xref="GI:2313285" /translation="MVAVKKNEDNKKLYDIDGQRTTIFMLLHVANKONKQEKQDET RKLYVGGELKLEVAENQSEFFTLLEAAKENISOKKMQTPRASKIFLKF" /complement(1817. 1918) /gene="HP0188" /complement(1817. 1918) /gene="HP0188" /note="hypothetical protein; identified by Genemark; putative" /codon_start=1 /transl_table=1 /product="H. pylori predicted coding region HP0188" /protein_id="AAd07266.1" /db_xref="GI:2313286" /translation="MKTTIKEIFQAEQSYIPNYQDYAMKDKNFRDL" 2075. 2608 /gene="HP0189" 2075. 2608 /gene="HP0189" /note="similar to PTD:882531 SP:P52082 GB:U00096 PID:1789376 percent identity: 43.14; identified by



QY	191	AlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspLysIleuArgPheLysVal	210
Db	3606	GTGGATTAATTTCCGTGTCATTATAGGGGGGGCCAAATTTGGGACAAATTAATTTTGATAC	3544
QY	211	GlyLysAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyLeu	230
Db	3546	GATTTAGACACGAATTTTATAGATTTAGTACGCTTTGCTTTTGGGGGGTGCCTTCAAA	3487
QY	231	ValSerHisAspPheAspArgTyrTrpPalaSerHisSerAlaHisAsnAlaThrArgIle	250
Db	3486	GCCAAAGAAACCTTTGAAAGCGTATTTGAGATTCGACCGCTATATCCCTGTTCAC--TTA	3430
QY	251	IleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspLysThrSer	270
Db	3429	CTAAGAAC-----	3422
QY	271	ArgHisAlaLeuLeuArgTyrArgGlyThrValGluGlnSerProLeuTyrGlnLysIle	290
Db	3420	--CATAAAGACACCAAA--AACCAAGCTAAAGAAATCGCTAACTCATGAAAAAATC	3367
QY	291	GlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeu-----	304
Db	3366	CCATACAGCGCTGAGACAAAAAACGATTGAAAAAAAAAGTCAATGATTTATAGATCGT	3307
QY	305	-----IleSerAspAspPro	309
Db	3306	TTCCAAAATACCAATACCCCATTTATATATGGAGATAGCCATTTTATTACCGATTCCACC	3247
QY	310	AlaLysGlyLeuAspArgAspArgGlySerProIleAlaGlyArgLeuGlnAspAla	329
Db	3246	AAAAAA--ATTGACAGCCCGCTGTATTCG--CCATACAAATCGCTTTTGAGAAACC	3193
QY	330	LeuLysGlnProGlyLysSerValTyrLeuValSerProTyrPheValProThrLysSer	349
Db	3132	CTTAAAAACGCTAAGACATCCGTTTTATATGCTTACATCGATTTTATTTCCAGGCAAAAAG	3133
QY	350	GlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspArgValLeuThrAsn	369
Db	3132	ATGATGAATAATCTTTAAATAATCAAAATTTCTAAAGGGATGAATGAACATCTTACCAAT	3073
QY	370	SerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysIleArgLysPro	389
Db	3072	TCCCTTCATCTACTGATGCGATAGTCGCTATAGGGGCAATGGCAAGATATGCAACCAA	3013
QY	390	LeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaValProAlaThr	409
Db	3012	TTAGCGCAATGGCGGAGATGTCCTATGAATAATGCAAAACGATTTTTCACACCCGCAAT	2955
QY	410	LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleVal	429
Db	2952	AAAGGCGC-----TTTAGACCAAAACATTCCTGCATGCAAGACAGATGTGTTTT	2902
QY	430	AspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsn	449
Db	2901	GATGACAAATTTAAGCTTTAGGGAGTTTCAAAATTTGATTCGCGCTGCAATACATCAAC	2842
QY	450	ThrGluMetGlyValValIleGluSerProLysIleAlaGluGlnMetGlnArgThrLeu	469
Db	2841	ACTGAAGCGGGTTTGTGTTGGACAAACCGCTCTTTTGCTTAAAGGGTGGCGTTTGTCGCT	2782
QY	470	AlaAspThrSerProGlnTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGln	489
Db	2781	AAAGATCATGCC--CAACATCATGCAATTTGCTGTGATGACGGAT--AGAGTGAT	2728
QY	490	Trp-----HisAspProAlaThrArgLysThrTyrPro	500
Db	2727	TGGGAAGCGGTGGAAGAGCATTTTATATCCATGAA-----AAAAT-----	2686
QY	501	AsnGlnProGlnAlaLysLeuTrpLysArgIleAlaLysIleLeuSerLeuLeuPro	520
Db	2685	---TCGCGTGCACCTTCTCTTTTGGCGGCTTATTAAGAATGCGTTAAAGTCTCTCT	2629

```

oy      521  IleguSerUen 524
Db      2628  GAAAGAGACTT 2617
RESULT  24
AC090533/c
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
COMMENT
-----Genome Center
Center: Harvard Partners Genome Center
Web site: http://www.hpgcg.org/Sequence/mouse.html
Contact: hpgc@emdel.mgh.harvard.edu
-----Summary Statistics
Center project name: ABU
Sequencing vector: pUC18, L08752
Chemistry: Dye-terminator Big-Dye, 1008
*Consensus quality: 257441 at least Q20
*Consensus quality: 253718 at least Q30
*Consensus quality: 247095 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 27165 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 6.8 x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 30267: contig of 30267 bp in length
* 30268 30287: gap of unknown length
* 30288 30288: contig of 28113 bp in length
* 58420 58400: contig of 28113 bp in length
* 58421 58420: gap of unknown length
* 88865 88864: contig of 30444 bp in length
* 88865 88884: gap of unknown length
* 109889 109889: contig of 21005 bp in length
* 109900 109900: gap of unknown length
* 109910 109910: gap of unknown length
* 125133 125133: contig of 15224 bp in length
* 125134 125133: gap of unknown length
* 125154 139549: contig of 14396 bp in length
* 139550 139549: gap of unknown length
* 139570 153410: contig of 13841 bp in length
* 153411 153430: gap of unknown length
* 153431 162567: gap of unknown length
* 153431 162567: contig of 9137 bp in length
* 162568 162587: gap of unknown length
* 162588 171044: contig of 8457 bp in length

```



[illegible]

TITLE	The genome sequence of the plant pathogen <i>Xylella fastidiosa</i> . The <i>Xylella fastidiosa</i> Consortium of the Organization for Nucleotide Sequencing and Analysis
JOURNAL	Nature 406 (6792), 151-157 (2000)
MEDLINE	20365717
PubMed	10910347
REFERENCE	2 (bases 1 to 10626)
AUTHORS	Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarengs,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carter,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorfy,H., Fachinca,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohne,M., Furlan,L.R., Garneri,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hobeisel,J.D., Junqueira,M.L., Kemp,E.L., Kitaajima,J.P., Krieger,J.E., Kurmae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Martins,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miranca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pessegueiro,J.B., Quaggio,R.B., Roberto,P., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawaeski,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,M.J., de Souza,A.A., de Souza,A.P., Terezi,M.F., Truffi,D., Tsai,S.M., Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Melandis,J. and Setubal,J.C.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil
FEATURES	location/Qualifiers
source	1. 10626 /organism="Xylella fastidiosa 9a5c" /db_xref="taxon:160492" /c1one="9a5c" 154. 795 /gene="XF1084" 154. 795 /note="similar to GI15381403 (percent identity: 32 %/query alignment coverage: 86.9 %/subject alignment coverage: 86.4 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=1 /product="partition protein" /protein_id="AAF3894.1" /db_xref="GI:9106037" /translation="MKREPVASGKGVGKTTIANTLAAGALRGVTVIADADPQKSS TRTERANDSOVLSTINANGANGSHMRNDLPNTLLIYDARGAFADLBEELEQ ADAIVPVMAALDIEAVGVGLNTMAVPRVHQKRLPVGVLNARSRTOTDQAMOM LGMDPYLVAQLRDSQFYVVLAGGRSVDFRSQAREHODNMKPLDHKL" 812. 1291 /gene="XF1085" 812. 1291 /gene="XF1085" /note="similar to GI15525061 (percent identity: 35 %/query alignment coverage: 91.8 %/subject alignment coverage: 84.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=1



```
gene
/produkt="conserved hypothetical protein"
/protein_id="AAF83895.1"
/db_xref="GI:9106038"
/translation="MMRELLIRHAHPGIGOTDEPRPLSOOGMTEAAGWMLCK
ORLIPDVLCSPARRTRETELEAYVAGDVQCEERLEATREACTLVALADECDYDR
LLVGNHPSMEQLSLHSNKTSDVGMPTGIALTLPLKTGIELDTICTALMWP
1311..1883
/gene="XF1086"
1311..1883
/gene="XF1086"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83896.1"
/db_xref="GI:9106039"
/translation="MSAMWIFILESVLPVSWAAQQLTLDVSHSYFEFEVTRTLCQ
SIDGVFPRLEGVVSLEPEGEKVLHLVAGDAVPGRTHTDWRGHEFPDAQHPPI
EFESQPYSPSRVLMLOGKISGNLIRGISRETLQIQPAACARPGYDCDVGTGVLRG
RYGIEDMRILGNVTFRMHGRLEGGPSH
1880..3838
/gene="XF1087"
1880..3838
/gene="XF1087"
/notes="similar to GI:4154695 (percent identity: 33 %/query
alignment coverage: 53.7 %/subject alignment coverage:
69.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/product="cardiolipin synthase"
/protein_id="AAF83897.1"
/db_xref="GI:9106040"
/translation="MNFIRLLLAALLLSACASLSHAQIDHAASVAVARPTLDC
DRPHKALDEPLVALGSRALSESTPVPLHVTLLDGESALVARIMLRASIDIM
QTYIFDKDDSRRLIMDELTAARGVRLIDQLSAISDLNIGALAGAVNSQIL
YNPIFEKAKLNGDYVASVLCERFNRQRMHMKLLVIDEMGVVGRNODDYDML
EYNERDVLVAGPAVLOMAVNFDAFMAAESVVECLRDYGRVLLDQVPLBSAVF
HPLERVORVTEANDPDYKRTFVDPALAVKQVYADLPQKHREHNANPVSIGPL
DSLISNARHEVLOTPTPLVSKPALNFRNLNRQDPRVYVYVNSLAIDNPVYAL
FYKRRNRMDLGDITETKPFEPONPIDLGLVPMGMNNDLSQTDLLAAKAVSD
DSDNAQLPSLENGQVDRVLRLETETRPPLGLHTVAKPLPVTBPGARMLHAASLV
DRKVGIGTHNFPREGESYNTAEVAVIEDAFAILLASLEGDDPCNANVVAARKKL
PGLYKLVSYGKLSEALPVLDMPEWYATVEERPGDCEPMLPRMDPSFMCYEPVG
DPEVNVGPKMLLYRMLIARGAGLPIL"
complement(4016..4225)
/gene="XF1088"
complement(4016..4225)
/gene="XF1088"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83898.1"
/db_xref="GI:9106041"
/translation="MWIGELYFSSQIKLTCPTITGQMONTLYLKYNHIIKKNKNS
RDRYLFERBSHPQFESVNLNFKM"
complement(4538..4699)
/gene="XF1089"
complement(4538..4699)
/gene="XF1089"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83899.1"
/db_xref="GI:9106042"
/translation="MLIASFLHCGKRPAYLAIFDDYISGVGDCIKLKADNAYTHKIYE
ILDKRIT"
complement(4845..5255)
gene
```

```
CDS
/gene="XF1090"
complement(4845..5255)
/gene="XF1090"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83900.1"
/db_xref="GI:9106043"
/translation="MYTLCRCSHLAMPKRNRSVAIEVGERLAARREAIPIRSK
DAAYLSKRLMKQVSHTTISDYESGSRPPPIVDALCOFYGTIPAAVVLGLMSRAA
VLAQKELSSSEERREIDRMLNMRISKTTD"
complement(5642..6010)
/gene="XF1091"
complement(5642..6010)
/gene="XF1091"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83901.1"
/db_xref="GI:9106044"
/translation="MPRLKSRNQLSRTSTSCQICFSQRIANATPNVQNAESV
NTEFLRICRREDTMRNCSNVASVRCPPHVPKRSRLRISASPHHAGTQAHRRHTQVNR
HTGQAHRRHTQVNAHSCSN"
complement(6090..6293)
/gene="XF1092"
complement(6090..6293)
/gene="XF1092"
/notes="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAF83902.1"
/gene
gene
CDS
Alignment Scores:
Pred. No.: 3,63e-23 Length: 10626
Score: 427.00 Matches: 158
Percent Similarity: 39.83% Conservative: 71
Best Local Similarity: 27.48% Mismatches: 202
Query Match: 15.74% Indels: 144
DB: 1 Gaps: 21
US-10-066-551-4 (1-525) x AEO003945 (1-10626)
QY 47 SerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGlnIleArgHis 66
Db 2060 AGCCGAGCGCTGAGTAATCTGACGCCCTGTG----- 2089
QY 67 ThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAspProHisGluAlaPhe 86
Db 2090 ACACCA-----CTTACATGATGACCATCTCGATMAAGCGAGCGCGCTTG 2137
QY 87 AlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIle 106
Db 2138 GTTCGACGAGCATATTGATTCGCGACGTGCTACCCGACATGATGATGACGACTTACATT 2197
QY 107 TyrArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleGluArg 126
Db 2198 TTTCGACGAGCATATTGATTCGCGACGTGATTAAGTATGAGCTTGTGACTGCTGTGAGCTG 2257
QY 127 GlyValArgValArgLeuLeuLeuAspAspAsnThrArgGlyLeuAspAspLeuLeu 146
Db 2258 GGTGTTCGAGTTCGTTTACTGATGATGACAGCTTTCGGCGATTTCGATCTGATATTCGTG 2317
QY 147 LeuAlaLeu---AspSerHisProAsnIleGluValArgLeuPheAsnProPheValLeu 165
Db 2318 GGTGCTTGGCTGGTGGACACGATTATTCACATGCGTATCTACACCGCATTTTGGC 2377
QY 166 ArgLysTyrPargAlaLeuGlyTyr-----LeuThrAspPhePro 178
```

```

Db 2378 AAA-----GCCAAGCTAAGTATGCGACATATGTTGGACAGCTGCTGTTGTTTGGT 2431
Oy ArgLeuAsnArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrLeu 198
Db 2432 CGCTTATACAGCGCATCACAATAATGTTGGTCATGATGAAATATGCGCGTGC 2491
Oy 199 GlyAlaArgAsnIleGlyAspLeuThrPheLysValGlyLysAspThrValPheAlaAsp 218
Db 2492 GCGCGGCGCTAATTACCAAGATGACTATATGATGGCATCTTGAATATACCTGCTGAC 2551
Oy 219 LeuAspIleLeuAlaThrGlySerValGlyLysLysSerHisAspPheAspArgLys 238
Db 2552 CGTATGATGCTGTGCGCGACCGCGCTGCTGCAATATGCGATTAACTTGATCGCTT 2611
Oy 239 TrpAlaSerHisSerAlaHisAsnAlaThrArgIleLeuArgSerGlyAsnIleGlyLys 258
Db 2612 TGGCG-----GCTGAGGCGAGCGCTACCGGTAGAGTGGCTCAGGAGTGGTGCAGC 2662
Oy 259 -----GlyLeuGlnAlaLeuGlyLysAsnAspLeuThrSerArgHisAla 273
Db 2663 ATGAGCTGACAGATGGGAGTGGCGCATGCTG-----CCATCAGCTGTCTTCCATCCA 2713
Oy 274 LeuLeuArg-----TyrArgGluThrValGlu-----GlnSerProLeuThrGlnLys----- 289
Db 2714 CTGGAACGGGTCACAGCGTGTACTGCTGACGCCAACAGATCCGAGCTATGTCACAGCGTACC 2773
Oy 290 -----IleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeu 304
Db 2774 TTGTGCATCTGCGCTGCTGCTGTCGACCAAGGACATAT----- 2812
Oy 305 IleSerAspProAlaLysGlyLeuAspArgAspArgLysPro-----ProIleAla 323
Db 2813 GTTGTGATCGCCGCCAAGAGATCCCATGACCAATGCAAAATCCGTTTCCATTTGG 2872
Oy 324 GlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyr 343
Db 2873 CCAGCGTGGATGATGTTGATTTCTAATGCGGCTCATGAGTGATCTTGCAGACCCCATAT 2932
Oy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeu-----ValGlnAspGly 361
Db 2933 CTGGGTATTCACAAACGACGCTGACATCTCCGCTCCATCGACACGACGACAA 2992
Oy 362 IleAspValThrValLeuThrAspSerLeuGlnAlaThrAspValAlaValHisSer 381
Db 2993 CCGCGTGGTGTGTCGACCAACAGTTGCTGCTACCGAATATCCATGCTATATCC 3052
Oy 382 GlyTyrValLysTyr-----ArgLysProLeuLeuLysAlaGlyLysLeuTyrGluLeu 400
Db 3053 TTATTTTACAAATACACAGCGGCGTAAATGCGTGCATGCGGTTGACATTTAGACAGTAC 3112
Oy 401 GlnPro----- 402
Db 3113 AAGCGGTTCCGCAAAATCCGCCATCGACCTGACGGCGTTGTCGATGATGATGATG 3172
Oy 402 ----- 402
Db 3173 AACAACTACTCTCTTACAGACGACAGACCTTTGGCTGCTGCCAAAGTCCGATGAT 3232
Oy 403 -----AsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerVal 419
Db 3233 TCAGCAATGCTGTCGATTTACCGTCTGGAATAAAGGTCAAGTGAATAGGCGTGTG 3292
Oy 419 ----- 419
Db 3293 TTACGTAAGAACGCTGACACCTTTTGGGCAATTCACACTGTCATTAACCATTTGCT 3352
Oy 420 -----ThrSerLeuHisAlaLysThrPheIleValAspGlyLys 432
Db 3353 GTGACTCGCCGCGGCGACGATGGCGTTGCATGCAAAATCATTTGGTCATGATGATG 3412
Oy 433 ArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAspThrGluMet 452

```

```

Db 3413 GTTGGGCTGATGATGCCCAACTGTCATGTCGGCGGTGACAGATTACATACTGAAGCT 3472
Oy 453 GlyValValIleGluSerProLysIleAlaGluMetGluArgThrLeu-----AlaAsp 471
Db 3473 GTTGGGCTGATGATGACATCAGCGCGTGGCTGCTGCGGCTTCACAGATGAGGTCAT 3552
Oy 472 ThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrHis 491
Db 3553 ATGATCCAGCTAATGCTGCGTGTG----- 3559
Oy 492 AspProAlaThrArgLysThrTyrProAsnGlnProGluAlaLysLeuThrLys----- 509
Db 3560 -----GCACCGCGGAGAAATTTACCTGCG-----CTCTATTAAGCTCAAT 3598
Oy 510 ArgIleAlaAlaLysIleLeuSerLeuLeuProIleGluSerLeu 524
Db 3599 TATTCGTAAGGTAAACTGTCGAGCATGCTGCTGCTGACCTG 3643
RESULT 26
LOCUS AE012113 11218 bp DNA linear BCT 23-MAY-2002
DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 21
ACCESSION AE012113 AE008922
VERSION AE012113.1 GI:21111127
KEYWORDS Xanthomonas campestris pv. campestris str. ATCC 33913.
SOURCE Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
ORGANISM Xanthomonas.
REFERENCE 1 (bases 1 to 11218)
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madelra, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spicola, L.A.F.,
Taki, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsal, S.M., White, F.F., Setubal, J.C. and
Katajima, J.P.
TITLE Comparison of the genomes of two Xanthomonas pathogens with
JOURNAL Nature 417 (6887), 459-463 (2002)
MEDLINE 22022145
PUBMED 12024217
REFERENCE 2 (bases 1 to 11218)
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,
Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardoso, J.,
Chambergo, F., Ciapina, L.P., Ciccarelli, R.M.B., Coutinho, L.L.,
Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S.,
Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
Madelra, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandis, J.,
Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
Okura, V.R., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spicola, L.A.F.,
Taki, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
Santos, M., Truffi, D., Tsal, S.M., White, F.F., Setubal, J.C. and
Katajima, J.P.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil

```

FEATURES		Location/Qualifiers
source	1..11218	/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
		/strain="ATCC 33913"
		/db_xref="ATCC:33913"
		/db_xref="taxon:190485"
gene		/note="pathovar: campestris"
		/complement(124..3696)
		/gene="lorA"
		/note="XCC0170; AF1489"
		/complement(124..3696)
CDS		/gene="lorA"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="indolepyruvate ferredoxin oxidoreductase chain alpha"
		/protein_id="AAM39489.1"
		/db_xref="GI:21111128"
		/translation="MRPLMOHLRDVAGLDITAGFVSGVRSPLGLDLELMRAHL EAKYTFPGNLDAITMWTGTOOTNLFPGATVGVGMWGRKPGVDRSGDVKHA NAAGTSPFGVALADADHACRSSTLPHGSEDEFVSAMPIINPAGVDILDMGLLGW AMSRTGRWVGFKIATFVESASVDVDFPFAQIVLPDDFALPGLSIRMPDPVDO EMRLHRYAVAAQAARANGIDKVLDSPOARLGIVTGKSYLDVQLAEYGLDQA CAQIGIRVYKGMTPLEPVGITAPARGLDIVVEKKAFIEROMKELEFYMPSAG ARPSIVKYNEHGAWILPSTGELPAPATIAVIGRKIORPPLPSAIAASIEGRLEME AKEMALPRLALFPVPHVYSCGCPHNTSTVDESGRAGLGCGCHYAMWMMNSTPTET HMGEGVTWSQAPPTATPHVFQNLGDTYHSGSLAIRQSATGVNITKILYNDAY DAVTGQPDVGLTVPDIAHQLEAGIEHETVAVSDIGKWRTRRRPESDVAHDSSEL DAVGQRLTIKGVSLILPDQCATGKRRRRKRLIDPEKRVMLNSLYCEGCGDEK SECVSVLPKDTFEGRRREIDQSNCKDYSCVGFSPFETVHGAPRKAGKADASAL DALPEPPORTLEOPMNLITGVSGTGVVITGALLGMAGHLEGKATVLDQGLKOG GAVTHIRIAROPADIAHVRIAGEADVLVGDWVVDVQALATALLGPAISNLEVIG EAMSGSTTODIOPRAEITAGVETALGDEPILLDAPQALATALLGPAISNLEVIG YAKHGVLPLSHALMRAIELNGAIVANMOQAFANGRIANDLPVQRAGLAPLDT YAAHAHAATPSPSGMEDHDSGQAPRALAQTDADTLHADGAPVHPLDEQLAOS LDEAIAARVHEFLTDQADAYAQOYRALVERVTHEQARPGSTALSAAVARAFKIMA YKDEYVARILYTRGDFORVOOQFEGDQLRFHLAPPLFAKDKANGOPIKRYGPMF SAFKLRLAKFLRGKLDVFGYTAERGEROLADYLVNTLRLDNDVNGLAQI ASIPHIRGYGHVKDHLHEAKARERALLAQRNPKALHYVQA"
gene		3996..4634
		/gene="XCC0171"
		3996..4634
CDS		/gene="XCC0171"
		/note="putative; ORF located using Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="conserved hypothetical protein"
		/protein_id="AAM39490.1"
		/db_xref="GI:21111129"
		/translation="MKROYAVYALALIGVWMAOQSLPKREFFEDERATTKPV VAVGQGDVYDRLAGVYQORDARAEPRAQIALALAYAGRTOLGDELYOGALALVNSG SQOTRAISMNTGMDLLRADKADKALQHWANLANGRPATPAMLPTTLALAGVAGKOE AEWYAAAVRTMPTDRWGSTANYASLLPWRREADRASLAEVFAWQANPAPF"
gene		4954..5871
		/gene="XCC0172"
		4954..5871
CDS		/gene="XCC0172"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="conserved hypothetical protein"
		/protein_id="AAM39491.1"
		/db_xref="GI:21111130"
		/translation="WAVDAFALILAMLGIGLVEARLVLPDNSADVLRNIVLYICLPA SVLTVPRHLHDSAGVYAPFWMLTALIVPMGCSRLGVRDEYALALMCVETN SSFIEPMVRAIIGDALPYAVVDQFGFVLSTFGVLVARYSGDPTPTAMRLAR VLRPPLMALVFLIMPEOPPAWIGSGKSLADAMLPVMLAVGESLQRLPADELK PLALGLFRIVAMPVIALPLSWALIGLOMLOTNVLESAMPMTITAAALAIISHRLAP LAAWVGSITLSLTLTPAWMLARLAA"
gene		5909..7003
		/gene="XCC0173"
		5909..7003
CDS		/gene="XCC0173"
		/note="putative; ORF located using Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="hypothetical protein"
		/protein_id="AAM39492.1"
		/db_xref="GI:21111131"
		/translation="NRALTGQTRMTRTSRLIGSYVLWLGSCNMAAAPAAPTADPT NDPLVQWQGVGETLQARLRSMADSGKPGPLYLAGLMLIDTETDAESDAAGVAPL ORAWLRALDARPDLLVARMALCPAGLRCDPLAALAEQADADAVLRAYAS AQRGDAAAEERAMQAAVQADHFDGALALGQALAAVADVOMPVLTYSLRQLDAR GLPSTGADPAAEFAMAAWSAHLPALSPVRCRDPATVTPALRDCMAVLGVAKDES TLATALICAKRWASISGADPTTRMORARPELAWLQOQOGLAGARGLVSDGAVLAH GEVPALRALQLOGTATOPPPAGMOPSAVQ"
gene		7363..7995
		/gene="para"
		/note="XCC0174"
		7363..7995
CDS		/gene="para"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="partition protein"
		/protein_id="AAM39493.1"
		/db_xref="GI:21111132"
		/translation="MKTIVLAKSGKGVKTTATNLAAHALQGRYLYADADPGSS TRVARRRSALSAVPLDTTRRRNQAAVPDSTDOLLIDAPGAADADISGLDHDVA ITVPVLSALDIEAVVGLNLTAKPVVQRLPVGLVLRNRPMTOTSOQAAMIGS WYPLVTLQDROQAVVMAAGLRSFLFYRSQVVRHODWPELLMKLKA"
gene		8011..8487
		/gene="XCC0175"
		8011..8487
CDS		/gene="XCC0175"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="conserved hypothetical protein"
		/protein_id="AAM39494.1"
		/db_xref="GI:21111133"
		/translation="MRELILRHAHAEPADTGOADPDRPLSPHGIAEAEAGRWLREQ HLPDPRVLCSPARRSRETEAVLELTGYEORLDRITYEATGTGLASYLDEHREARL LILGHNPGLERLALMLHSGQTDYGMPTASIALALPLDAIEPGIARLTAFWMP"
gene		8491..9063
		/gene="XCC0176"
		8491..9063
CDS		/gene="XCC0176"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="conserved hypothetical protein"
		/protein_id="AAM39495.1"
		/db_xref="GI:21111134"
		/translation="MPLSLRWAMLAWLPCVAQOKVHAIDPQSRGFEIRTRFG MKIEGLFRFRGELTESADGHQVRRFLDAADVLEIPGDKRTYAMRGDFEFVAQYV VFESFPYSAEVVKKOGETTGNLTIRGITHVETLHVPAECARPGTDCDVISRGTVLR GRYGNAWQMALGDVYTLIRGRLOEVRRP"
gene		9060..11042
		/gene="cis"
		/note="XCC0177"
		9060..11042
CDS		/gene="cis"
		/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
		/codon_start=1
		/transl_table=11
		/product="cardiolipin synthase"
		/protein_id="AAM39496.1"







Db 5034 GACCCGCGCGAGCAAACTACACACCGAGCGCGCTGTGATGCAGCATCGCGCTTT 5093  
 QY 463 AAGlUGlmeGluArgThrLeu---AlaAspThrSerProGluTyrTrpAlaTyrVal 480  
 Db 5094 GCGACGCGCTGTGCTGCCACATCGACACCGCATCGCGCAAAATCGTGGACGCTG 5153

# RESULT 28 LMFCHR16\_00 WPCOMMENT

Sequence split into 11 fragments LOCUS LMFCHR16 Accession ALA99619

Fragment Name	Begin	End
LMFCHR16_00	1	110000
LMFCHR16_01	100001	210000
LMFCHR16_02	200001	310000
LMFCHR16_03	300001	410000
LMFCHR16_04	400001	510000
LMFCHR16_05	500001	610000
LMFCHR16_06	600001	710000
LMFCHR16_07	700001	810000
LMFCHR16_08	800001	910000
LMFCHR16_09	900001	1010000
LMFCHR16_10	1000001	1030105

DEFINITION LMFCHR16 Leishmania major chromosome 16,17 clone Chr.16,Chr.17 strain  
 ALA99619  
 FIREDIN, \*\*\* SEQUENCING IN PROGRESS \*\*\*; in unordered pieces.

ACCESSION ALA99619.1 GI:11877282  
 VERSION HTG: HTGS\_PHASE1.  
 KEYWORDS Leishmania major.  
 SOURCE Leishmania major.  
 ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;  
 Leishmania.

REFERENCE 1 (bases 1 to 1030105)  
 AUTHORS Murphy,L., Quail,M., Harris,D., Rajadream,M., Ivens,A. and  
 TITLE Direct Submission  
 JOURNAL Submitted (15-DEC-2000) Major Genome Sequencing Consortium, The  
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge  
 CB10 1SA, UK

COMMENT For more information about this sequence or the Leishmania project,  
 see <http://www.sanger.ac.uk/Projects/Lmajor/>  
 CAVEATS:  
 1. the sequencing is still in progress  
 2. this submission represents 800 unordered contigs of size 0.5kb  
 or greater, separated by blocks of N  
 3. it is derived from whole chromosome shotgun plus cosmid and/or  
 PAC skims  
 4. the data are EXTREMELY preliminary  
 5. the sequence may contain E. coli, sequencing/cloning vector, or  
 be cross-contaminated with other Leishmania chromosomes IMPORTANT:  
 This sequence is unfinished and does not necessarily represent the  
 correct sequence. Work on the sequence is in progress and the  
 release of these data is based on the understanding that the  
 sequence WILL change as work continues.  
 THE ORDER OF CONTIGS IS NOT KNOWN.  
 THESE DATA ARE PRELIMINARY ONLY, and annotation will follow  
 shortly.

\* NOTE: This is a 'working draft' sequence.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

## FEATURES source

Location/Qualifiers  
 1..1030105  
 /organism="Leishmania major"  
 /strain="Friedlin"  
 /db\_xref="taxon:5664"  
 /chromosome="16,17"  
 /clone="Chr.16,Chr.17"

BASE COUNT 185925 a 289649 c 284080 g 190256 t 80195 others  
 ORIGIN

Alignment Scores:

Pred. No.:	9,81e-15	Length:	110000
Score:	336.00	Matches:	122
Percent Similarity:	42.41%	Conservative:	68
Best Local Similarity:	27.23%	Mismatches:	161
Query Match:	12.38%	Indels:	98
DB:	2	Gaps:	16

US-10-066-551-4 (1-525) x LMFCHR16\_00 (1-110000)

QY 168 TTPATGAlaLeuGlyTyrLeuThrAspPheProAlaLeuAsnArgArgMetHisAsnLys 187  
 Db 3975 TGCGCGCGCTGTGATGATACCGACGAGAT-----GAACACCGCTCTGTGATCATCGA 4025  
 QY 188 SerPheThr-AlaAspAsnArgAlaThrIleLeuGlyArgAsnIleGlyAspGlu-- 206  
 Db 4026 AGCGGGCGATCGCGATGCGCCCAAGCTGTGTGGCGGCGGCAACGACGATGAGGAAAG 4085  
 QY 207 -----TyrPheLysValG1 211  
 Db 4086 CGGTGGCACCATTATGTCAGACCAAGATTTTCAGCGCGTGCACCAAGCCATGAAGTCGC 4145  
 QY 211 yGluAspThrValPheAlaAsp---LeuAspIleLeuAlaThrGlySerValValG1G1 230  
 Db 4146 CCAAGAGAAATCTTGGCCCGGTGCTGTGATGCTGTCGACAGCGGTGAAGAAAGC 4205  
 QY 230 uValSerHisAspPheAsp-----ArgTyrTrpAlaSerHis-- 242  
 Db 4206 CATTCGATCGCCACAGACACCAATACGGCCGTGGCGCGGTGTGACGTCGAACAT 4265  
 QY 243 ---SerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGly----- 257  
 Db 4266 CTCACAGATTTTCTACCGCCCAAGCAGCAGCTGCGCGGATGCGGTGACCAACAGTA 4325  
 QY 258 -----LysGlyLeuAlaLeuGlyTyrTrpAsnAspG1 268  
 Db 4326 CGATGGCGGCGATGACTGACACCGTTCGCTGTTTCAACAGCTGCGGTAACCGCGCA 4385  
 QY 268 uThrSerArgHisAlaLeuLeuArgTyrArgGlu-----ThrValGluGlnSerPr 285  
 Db 4386 CAAGTCGCTGCACCGCTTCGACAGATACACCGAGCTGAAAGCAGCAGCTG- GATCAAGTGT 4444  
 QY 285 oLeuTyrGlnLys----- 289  
 Db 4445 AAGCTTCAGAAACACCGC----- 4504  
 QY 289 ----- 289  
 Db 4505 NNT 4564  
 QY 290 ---IleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAspAs 308  
 Db 4565 GGGCATCTGCGCGCGGAGACTGATCTGG-----GGCTGAACACGAGCTGGGATGC 4618  
 QY 308 pProAlaLysGlyLeuAspArgAspArgArgLysProPro-----IleAlaGlyArgLe 326  
 Db 4619 GCGGACCAAGGTGCTGCGCAAGCGCATCCGAGTCCAGCGCTGCTGACCAACCGAGCT 4678  
 QY 326 uGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProGlyPheValPr 346  
 Db 4679 GGCACCGGAGCTGGAAGGGGTGAGCCACGAGTGTGATGATGCTGACTTACTCTGCC 4738  
 QY 346 oThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrVa 366  
 Db 4739 GGGCCAGCTGCGCTGTGATGACTGACCGGTGCGGAGCGCGGTCTCGGTAGCCCT 4798  
 QY 366 lleuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTy 386  
 Db 4799 GGTGACCAACTCTCTGGAAGCTACCGAGCTGCGCGCGGTGATGCGGCTA 4858  
 QY 386 rArgLysProLeuLeuLysAlaGlyTyrLeuLysLeuTyrGluLeu-----GlnProAsnH 404  
 Db 4859 CCGCAAGGATTCGTGAGACGAGAGTGAATGTAAGAGTTGCGCGCGCGCGGAGCA 4918





LOCUS BUSERP 10118 bp DNA linear BCT 19-MAY-1997  
 DEFINITION B.subtilis urease operon and downstream DNA.  
 ACCESSION Y08559  
 VERSION Y08559.1 GI:1592693  
 KEYWORDS urease  
 SOURCE Bacillus subtilis.  
 ORGANISM Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 REFERENCE 1 (bases 1 to 10118)  
 AUTHORS Cruz-Ramos, H., Glaser, P., Gray, L. V. Jr. and Fisher, S. H.  
 TITLE The Bacillus subtilis ureABC operon  
 JOURNAL J. Bacteriol. 179 (10), 3371-3373 (1997)  
 MEDLINE 97294484  
 PUBMED 9150240  
 REFERENCE 2 (bases 1 to 10118)  
 AUTHORS Glaser, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-SEP-1996) Philippe Glaser, Regulation de l'Expression Genetique, Institut, Pasteur, 28 Rue du Dr Roux, Paris, Paris, 75724, France

FEATURES  
 source 1..10118  
 /organism="Bacillus subtilis"  
 /strain="168"  
 /db\_xref="taxon:1423"  
 join(145..151,159..476)  
 /gene="urea"  
 145..151  
 /gene="urea"  
 159..476  
 /gene="urea"  
 /codon\_start=1  
 /transl\_table=11  
 /product="urease"  
 /protein\_id="CA69857.1"  
 /db\_xref="GI:1592694"  
 /db\_xref="SWISS-PROT:P75030"  
 /translation="MKLTPVDEKLLIFPAAGELAKORRARGVLLNPPAAAYITCFIM  
 EGARDGKGVAAELMEAGRHVLTKEVDMEGVPEKLDSIQEATFPDGVKLVTVHQPISAE  
 VKS"  
 461..467  
 /gene="ureb"  
 461..467  
 /gene="ureb"  
 473..847  
 /gene="ureb"  
 473..847  
 /gene="ureb"  
 /codon\_start=1  
 /transl\_table=11  
 /product="urease"  
 /protein\_id="CA69858.1"  
 /db\_xref="GI:1592695"  
 /db\_xref="SWISS-PROT:P71035"  
 /translation="MKPGAFQIABEGTITINGREIREVTYKNTGSRSLQVGSHPFEAF  
 ANGALLFEDRELAIQMRIDVPSGTSVRFPGSQKTVSLVEIGRKTIIRGLNGMADTFID  
 ERGRKEXTLANKQAGMEGVIR"  
 829..836  
 /gene="urec"  
 829..836  
 /gene="urec"  
 844..2553  
 /gene="urec"  
 844..2553  
 /gene="urec"  
 /codon\_start=1  
 /transl\_table=11  
 /product="urease"  
 /protein\_id="CA69859.1"  
 /db\_xref="GI:1592696"  
 /db\_xref="SWISS-PROT:P77837"  
 /translation="MKMSREYIAELFEPTTQDKIRLGDITLMIIEVEKDFTYGEBMIT  
 GGGKTIIDGKQNGRITGKDGADLVLTITNVLLDTYGLIVKADVGKQGRIVGKSGN

gene  
 RBS  
 gene  
 CDS

gene 2674..3089  
 /gene="ywna"  
 2674..3089  
 /gene="ywna"  
 2688..3089  
 /gene="ywna"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Unknown"  
 /protein\_id="CA69860.1"  
 /db\_xref="GI:1592697"  
 /db\_xref="SPTREMBL:P71036"  
 /translation="MINSRLVAIHLSLSISNDEKTSSEITINDSVNTNPPVYRRMISL  
 LKADILTSRAGVPGASLKKDPADISLLEVYRAVDQBELFVHNPKNPCVGRKIQ  
 NALDEFESVQRAMENELASKSLKDVNHLF"  
 3141..3798  
 /gene="ywnb"  
 3141..3798  
 /gene="ywnb"  
 3157..3798  
 /gene="ywnb"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Unknown"  
 /protein\_id="CA69861.1"  
 /db\_xref="GI:1592698"  
 /db\_xref="SPTREMBL:P71037"  
 /translation="MKIGIGASGRAGNEILKEARKRGHEVTAIYRNASKVOEDPAI  
 LEKDVFELEAETRIKPDVAVNAGAPGGEHLVAGRALISILMDAKRTLRLLVGA  
 GSLFVDEAKTETRLMDTPPERKXYLPTASOGENLKDLOOTDSISMTLSRAFPDPAG  
 KRTGSYOKCKDNVIYNAKDSIYSTADVAIVLDELHEHEHNERFTYVSEAE"  
 3800..3828  
 complement(3833..4233)  
 /gene="ywnC"  
 complement(3833..4216)  
 /gene="ywnC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Unknown"  
 /protein\_id="CA69862.1"  
 /db\_xref="GI:1592699"  
 /db\_xref="SPTREMBL:P71038"  
 /translation="MNPENMKTLISISWCKATGILFTMGATLALSAPFELIGAV  
 PGVLIITISGIFLMRSAREGMAEHNSGSEDMLENTAKFPKMGITLVLSIAVSIL  
 AIAAFIFLMDIADGLFSIDYSTY"  
 complement(4226..4233)  
 /gene="ywnC"  
 4347..5134  
 /gene="ywnd"  
 4347..4354  
 /gene="ywnd"  
 4361..5134  
 /gene="ywnd"  
 /note="Product similar to Streptomyces lividans  
 transcriptional activator TipA"

RBS  
 gene  
 RBS  
 CDS

terminator  
 gene  
 CDS





gene  
CDS  
  
SAFEQSLARLAEFGSGIHNCVGAFAFKNEIEIVANVLDMKRNIRLEEDFCYAESGLYALP  
RGPVSIAAFPODA"  
complement(5791. .6537)  
/gene="yvmc"  
complement(5791. .6537)  
/gene="yvmc"  
/function="unknown"  
/note="similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15524.1"  
/db\_xref="GI:2636033"  
/db\_xref="SPTREMBL:O34351"  
/translation="MTGVTERRSVFIAEALTENCREFERRRHVLVLSIPFSRPRES  
EDYIRILGMAOKFSSVSLAGHEANLLEALGTGPGKAERKVRKEVSNRRPAR  
ALVAIRGGMKAHTEPSDFIDNKAYOLROEVENHAFPEQDHPFRACLDMSRAITGAR  
CSLMERSSEDMNLAVEFYAELPFETIGAPDLVEETLLATHRPWKAGEKISNHE  
PSICHRPNQGITLVDEMAOMISEKRITSEG"  
6963. .7472  
/gene="yymb"  
6963. .7472  
/gene="yymb"  
/function="unknown"  
/note="alternate gene name: yzha  
similar to hypothetical proteins"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15525.1"  
/db\_xref="GI:2636034"  
/db\_xref="SWISS-PROT:P40762"  
/translation="MSDLTKRMIDYIYRLRLHLEOKANTSLOOFEFKEAAEDYAELP  
KNMTHIHVIDCGOHEPIINNGLARKMLNSAVNTKISTKLKEEFINSQTLDNKKRE  
VYEKTRRGRIREFDLHERLKRELAEYQFLDSFSOEOKAVLKFLEQLTSLEAQDT  
DETDPKPVK"  
7493. .8726  
/gene="yyma"  
7493. .8704  
/gene="yyma"  
/function="unknown"  
/note="similar to transporter"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15526.1"  
/db\_xref="GI:2636035"  
/db\_xref="SPTREMBL:O34307"  
/translation="WAHTKSNAVLILLTVCFSAFPASLSONITYSPILPIIKESHVST  
AMVNLSVEPMIVTAIMQILIGAITDEGARIVITGTLATPAASIGCAVTTFTEL  
IFRMIOAGSALPIATAFTTGOLFQFTGNERSMGATQMILSVAPAIPAVALGGTIGA  
AGYESIEMAIASIVLAVTNISITEPKDSPESMOAKGANVFAHKSIFFNRGNVL  
LTSLFEIYFIVEYIYLPILLETENHDIVDAGLLYLPLASTAGTFFLRKIQAOKIG  
LITLFIGNSYAACSIILFEAVTHSVSIYAMLATLALGISMCVTPPLXSMITNREP  
NRGSIGCNFNRTIRYGMAAGPWNSYLLTMPSAMSFSLLGLGFALSFCCLPPMFS  
DKRKOKHHM"  
8709. .8726

terminator

gene  
CDS  
  
/gene="yyma"  
complement(8732. .9091)  
/gene="yvld"  
complement(8732. .9091)  
/gene="yvld"  
/function="unknown"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAB15527.1"  
/db\_xref="GI:2636036"  
/db\_xref="SPTREMBL:O34648"  
/translation="WKAWSIILYNALLLYIDGYIDSIIHSIGAAIINSLISLIIN  
VLIKRLIIFLPLVTMTGLGLEFINAITLMTMASIMGDSFOIDGFIATMSVIIS  
VFHLIQGIEPLEPRKK"  
complement(9093. .9290)  
/gene="yvlc"  
complement(9093. .9290)  
/gene="yvlc"  
/function="unknown"

gene  
CDS

		/codon_start=1		
		/transl_table=11		
		/protein_id="CAB15528.1"		
		/db_xref="gi:2636037"		
		/db_xref="s:SPRMBL:034719"		
		/translation="MNKLYSEKNKKIAGVIGLAEYFNMDSALRFIVTILAMTSV		
Alignment Scores:				
Pred. No.:	1.67e-13	Length:	212610	
Score:	325.00	Matches:	124	
Percent Similarity:	38.12%	Conservative:	67	
Best Local Similarity:	24.75%	Mismatch:	186	
Query Match:	11.98%	Indels:	124	
DB:	1	Gaps:	15	
US-10-066-551-4 (1-525) x BSUB0019 (1-212610)				
QY	43	GIUATGTHGJLSEARFGRHISPHESNTHSRISLYSPROVALARGLEUASPASN-----	60	
		.      :::     ::  ::		
Db	164874	GAAAGATTTAGAAACCAACCAACCAATTTCCAGTTTAACAACCGGCAACCTTGATACCAAGAT	164933	
QY	61	-----IleuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIle	76	
		:::		
Db	164934	TTAATTATATGCTCATTTATGATATACCATGCTGCTGTTACAGAGAT--AATTCAAGT	164990	
QY	77	TYTLEULEUASPAPROHISGLUALAPHEALALALATARGALALALEUILEGJLSEVALA	96	
		:::     ::  ::		
Db	164991	GACGCATTTACAGATGAGCAGGATTAATTTACAGCGGCTGTGAGTGCATTTCAAGGCT	165050	
QY	97	GIUHSISERLEUASPARGGLNLTYYTLYTTLTPATGASNASPILSESERGLYARGLEU	116	
		::: ::       :::     ::      ::		
Db	165051	AAGGATCATATCCATTTACAGTATATATATATAAGCGATGATTTAGCCAAAATAATG	165110	
QY	117	PHESNLEUVALTYTLEUALALEGLARGGLYVALARGVALARGLEULEUASP	136	
		:::     ::  ::		
Db	165111	CGGAGCGCTCTCATCCAAAGAGCGAAGAGAGGTGTTCCAGTCAGTCTTTATGATGAA	165170	
QY	137	ASNAANTHARCTGGLYLEUASPAPSEULEUVALALASPERSERHISPROASNILEGLU	156	
		:::     ::  ::		
Db	165171	CTCGGGTCCAGACGCTGAGAAAAAGTTTTTAAAGAGCTTCGAGAGCTGGGGCGAT	165230	
QY	157	VALARGLEUPHEASNPROPHHEVALLEUARGLYSTRPATGALALEUGLYTYTLEUTHRASP	176	
		::		
Db	165231	GTGGAGGCTTTTCCCTCAAGCTCGAGCGCATTAACCTG-----	165272	
QY	177	PHEROPALAEUASPARGMETHISNLSNLYSERPHEPTRHIALASPASPARIALATHR	196	
		:::		
Db	165273	-----CGGCTGAATTTACCGGAAACCATGGAAGCTTGATCATAGACGGATACCGGT	165326	
QY	197	ILEUGLGLYARGASNILEGLYASPGILUYYRHE-----LYSVALIGLYLU	212	
		:::           ::  ::		
Db	165327	TATGTCGGAGGTTTACAGCTGGAGATGATGATAGCGCTTAATCCGAATAATGGCT---	165383	
QY	213	ASPHTRVALPHEALASPSEUASPILLEUVALATHRGILYSERVALVALIGLYLUVALSER	232	
		:::                 ::  ::		
Db	165384	-----TATTGGCGGGAATCCATATCCGGCTGCAGAGGAGCGGCTGCACGCCATCCAA	165437	
QY	233	HISASPERASPARGTYTTPALASERHISSERIALHISASNALATHARGILELEARG	252	
Db	165438	ACACGATTCATCTTGACTGGAACCAACCTTCTCATCAACCTACGCTGACATTAATTCCA	165497	
QY	253	SER-----GLYSNILEGLYLSGILYLEUGLNALALEUGLY	264	
		:::                 ::  ::		
Db	165498	AATCACTTCTCGATTTACGGGCCAAAAGCAACGTCGGAGATGACGATTTGCACAAAGGCGC	165557	
QY	265	TYRASNASPGILUTHRSEARGHISALALEUVALTYTARGILUTHRVALIGLULINSER	284	
		:::		
Db	165558	CCTGATTTGGAA-----	165569	
QY	285	PROLEUTYRGLNLYSILEGLDTHRGILYARGTILEASPTRGILINSEVALGINTHRARGLEU	304	
		:::		
Db	165570	-----TGGGAACAATTTAAAATGGCTACATC-----AAATGG	165602	

```

QY 305 IIESeRspAspPrOalAlaLysGlyLeuAspArgpArgLysProIleAlaGly 324
      |||||
      ATTCGAAC-----
DB 165603 ATTCGAAC----- 165611
QY 325 ArgLeuGlnAspAlaLeuLysGlnProGluYSerValTyrLeuValSerProTyrPhe 344
      |||||
      -GCCAAGCGTTCCATTGATACAAACACCTTATTTT 165647
DB 165612 -----
QY 345 ValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364
      |||||
      ATCCCAAGTCAAGCGTCTGGATGCTTTAAGATCGCTTGCCTTCGGCATATGACGTT 165707
DB 165648 ATCCCAAGTCAAGCGTCTGGATGCTTTAAGATCGCTTGCCTTCGGCATATGACGTT 165707
QY 365 ThrValLeuThrAspSerLeuGlnAlaThrAspValAlaAlaHisSerGlyTyrVal 384
      |||||
      AACATTATGATTCGCAAC-----AAGCCGATACAGCCTTGTATCTGGGAGACATTT 165761
DB 165708 AACATTATGATTCGCAAC-----AAGCCGATACAGCCTTGTATCTGGGAGACATTT 165761
QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHis 404
      |||||
      TCGTATATCGGCGATCTGCTGAAGCGGAGCCACTGCTAT----- 165803
DB 165762 TCGTATATCGGCGATCTGCTGAAGCGGAGCCACTGCTAT----- 165803
QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424
      |||||
      ATTTACGATTAACGCGTT-----ATTCACGCG 165830
DB 165804 -----ATTTACGATTAACGCGTT-----ATTCACGCG 165830
QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444
      |||||
      AAAACCATGCTGCTGATACCAAAATCCGTCAGTCGACACGCAACATGATGTCGCA 165890
DB 165831 AAAACCATGCTGCTGATACCAAAATCCGTCAGTCGACACGCAACATGATGTCGCA 165890
QY 445 SerAlaArgLeuAsnThrGlnMetGlyValIleGluSerProLysIleAlaGln 464
      |||||
      ACCTTCAGGCTGAATTTGAGGTAAATGCAATTTATTTACGATATCCCAATCTCTAAAAA 165950
DB 165891 ACCTTCAGGCTGAATTTGAGGTAAATGCAATTTATTTACGATATCCCAATCTCTAAAAA 165950
QY 465 MetGluArgThrLeuAlaAsp-----ThrSerProLysTyrAlaTyrArgValThr 481
      |||||
      CTGGTTTCAACTTTAAGAATATTTGCTAGTATCCAGAAATTCACCTATAGAGAGTAT 166010
DB 165951 CTGGTTTCAACTTTAAGAATATTTGCTAGTATCCAGAAATTCACCTATAGAGAGTAT 166010
QY 482 LeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLysThrTyrProAsn 501
      |||||
      CTGTCAGCG----- 166019
DB 166011 CTGTCAGCG----- 166019
QY 502 GluProGluAlaLysLeuTyrLysArgIleAlaIleLysIleLeuSerLeu-----Pro 520
      |||||
      -----CCGTTGTGATTCGAAATTAAGAAATCCGTCGCGCGCTGTGTGTCGCG 166067
DB 166020 -----CCGTTGTGATTCGAAATTAAGAAATCCGTCGCGCGCTGTGTGTCGCG 166067
QY 521 Ile 521
      |||
      ATC 166068 ATC 166070
DB 166068 ATC 166070

```

```

RESULT 32
AX414863
LOCUS AX414863 2319 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1854 from Patent WO0228891.
ACCESSION AX414863
VERSION AX414863.1 GI:21447320
KEYWORDS
SOURCE
ORANISM Listeria monocytogenes ATCC 19115.
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1 Glaser, P. and Kunst, F.
TITLE Listeria innocua, genome and applications
JOURNAL Patent: WO 0228891-A 1854 11-APR-2002;
Pasteur Institut (FR)
FEATURES
source
1. 2319
Location/Qualifiers
/organism="Listeria monocytogenes ATCC 19115"
/db_xref="Caxon:176281"
BASE COUNT 759 a 368 c 489 g 702 t 1 others
ORIGIN

```

Alignment Scores: 6.91e-16 Length: 2319  
Pred. No.: 323.00 Matches: 124  
Score: 323.00

```

Percent Similarity: 37.93% Conservative: 74
Best Local Similarity: 23.75% Mismatches: 214
Query Match: 11.91% Indels: 110
DB: 6 Gaps: 17
US-10-066-551-4 (1-525) x AX414863 (1-2319)
QY 4 AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23
      |||||
      AATCCAGCAACTGAAATATTC-----AGCACAGCTCAAGTATGAGAAAAGCTAAATTA 608
DB 555 AATCCAGCAACTGAAATATTC-----AGCACAGCTCAAGTATGAGAAAAGCTAAATTA 608
QY 24 IIESeRLeuLysLeuLeuLysSerCysSerSerTyrLeuProLeuGlnGlu 43
      |||||
      ATTAATGCATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTA 668
DB 609 ATTAATGCATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTCACCAATTA 668
QY 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
      |||||
      AGCAAT-----GCGGATTTAAGCTCATTAGAGCCCAATTAAG--GCGAATTAATAACGAA 719
DB 669 AGCAAT-----GCGGATTTAAGCTCATTAGAGCCCAATTAAG--GCGAATTAATAACGAA 719
QY 64 IleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuAsnAspProHis 83
      |||||
      ATATTA-----ACAAATGCG-----GAA 737
DB 720 ATATTA-----ACAAATGCG-----GAA 737
QY 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGln 103
      |||||
      GAAACTTTTCCAGTCTGCTTACACGCGCTTACGAAAGCTGAAACACCATTCATATTCAA 797
DB 738 GAAACTTTTCCAGTCTGCTTACACGCGCTTACGAAAGCTGAAACACCATTCATATTCAA 797
QY 104 TyrThrIleTyrArgHisPheAsnIleSerGlyArgLeuLeuPheAsnValTyrLeuAla 123
      |||||
      TACTATATTTTCAAAACGAGATGAAATTTCTACAAAGATTCGTGATTTTGTGAGAAAAA 857
DB 798 TACTATATTTTCAAAACGAGATGAAATTTCTACAAAGATTCGTGATTTTGTGAGAAAAA 857
QY 124 AlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArgGlyLeuAsp 143
      |||||
      GCAAAAGCAGCGGCTGAAGTATGATTATGTTGATGGGCTTGATGCAAGCAACGCGGT 917
DB 858 GCAAAAGCAGCGGCTGAAGTATGATTATGTTGATGGGCTTGATGCAAGCAACGCGGT 917
QY 144 AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163
      |||||
      AAAGCTTTTGTAGCTCTTGTGAAGAGAGCTGAGTTAGTATTCACGATTTGACCA--- 974
DB 918 AAAGCTTTTGTAGCTCTTGTGAAGAGAGCTGAGTTAGTATTCACGATTTGACCA--- 974
QY 164 ValLeuArgLysTyrPheArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArg 183
      |||||
      ATTACCTCTCCCTGGATGTA-----AGAACACCAATTTTAACA 1013
DB 975 ATTACCTCTCCCTGGATGTA-----AGAACACCAATTTTAACA 1013
QY 184 MetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyArgAsnIle 203
      |||||
      AATCACCCTAATAATTCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1073
DB 1014 AATCACCCTAATAATTCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1073
QY 204 GlyAspGluTyr-----PheLysValGlyLysPheValPheAla 217
      |||||
      GGAGAGAAATATCTCTATATACACCAAGATTTCCGTTTGGCGCATACCA----- 1124
DB 1074 GGAGAGAAATATCTCTATATACACCAAGATTTCCGTTTGGCGCATACCA----- 1124
QY 218 AspLeuAspIleLeuAlaThrGlySerValIleGlyLysValSerHisAspPheArg 237
      |||||
      -----CATATTAATAATTCACAGGCAAGCTGTATGCACTTCACAAATCTTTTGAAT 1178
DB 1125 -----CATATTAATAATTCACAGGCAAGCTGTATGCACTTCACAAATCTTTTGAAT 1178
QY 238 TyrTyrPala-----SerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsn 255
      |||||
      GATGGGCTATATGAGAAACCAAGCAGGTCGCGAGATCAATTTATGAGAGCTGGA 1238
DB 1179 GATGGGCTATATGAGAAACCAAGCAGGTCGCGAGATCAATTTATGAGAGCTGGA 1238
QY 256 IleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 275
      |||||
      TTACACATATATTTTCCACAGTGTGATGAGGAGCAATGAGCG----- 1283
DB 1239 TTACACATATATTTTCCACAGTGTGATGAGGAGCAATGAGCG----- 1283
QY 276 ArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleIleThrGlyArgIle 295
      |||||
      -----CAAGTGAATTTCAGCGGACCATAT 1307
DB 1284 -----CAAGTGAATTTCAGCGGACCATAT 1307
QY 296 AspTyrGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315
      |||
      GAT----- 1310
DB 1308 GAT----- 1310
QY 316 AspArgArgLysProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLys 335
      |||||

```

Db 1311 -----AAGAAAGAAAGGCTCGCGATTCGATTCGTTGATTAAATTCGCTTAAGAA 1364  
Qy 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355  
Db 1365 TCTGTTGGATTGTGTCCTTACTTGTTCAGATGAGAGATCCCTTCGATTCGCTTATTCG 1424  
Qy 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375  
Db 1425 CGGCTTCGATGAGCGGCTGTCAGCTCGAGTTATTATTCAGGC-----AAGAGTAT 1478  
Qy 376 ValAlaValAlaHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIle 395  
Db 1479 CGCGGATTTTCATTCACGAGATGATAGCTGATGTGAACAATGATGAGAGCGCG 1538  
Qy 396 LysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThr 415  
Db 1539 AAAATGATGCTTATGCGGATGATCTTTGTT----- 1571  
Qy 416 GlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePhe 435  
Db 1572 -----CATCGAAGCAATGCTAGTGGAGCGGAGCGCTCAGCT 1610  
Qy 436 IleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValAla 455  
Db 1611 ATAGTACTGCCAATTTTGACGCTGCTAGTTTACATTGATCAATGATTAATGATATTC 1670  
Qy 456 IleGlySerProLysIleAlaGluGlnMetGluArgThrLeuAlaAspThrSerProGlu 475  
Db 1671 TTATGTGATGAA-----AGCGAGCTATGCTATTTAAACGCTGATTTGAGAAAGT 1724  
Qy 476 TyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrHisAspProAlaThr 495  
Db 1725 TTT-----GAGATAGCGGACTT-----TTTACG 1748  
Qy 496 ArgLysThrTyrProAsnGluProGluAlaLysLeuTyrLysArgIleAlaLysIle 515  
Db 1749 ATGAAGATATGAAACAAACCATTTATGACTGCTATTAAGAAAGATCTTACTGTTTA 1808  
Qy 516 LeuSer 517  
Db 1809 CTATCA 1814

RESULT 33  
AX416851 6273 bp DNA linear PAT 14-JUN-2002  
LOCUS AX416851 Sequence 3842 from Patent WO0228891.  
DEFINITION AX416851  
ACCESSION AX416851  
VERSION AX416851.1 GI:21449461  
KEYWORDS  
SOURCE Listeria monocytogenes ATCC 19115.  
ORGANISM Listeria monocytogenes ATCC 19115  
REFERENCE 1 Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
AUTHORS Glaser, P. and Kunst, F.  
TITLE Listeria innocua, genome and applications  
JOURNML Patent: WO 0228891-A 3842 11-APR-2002;  
Pasteur Institut (FR)  
FEATURES  
source Location/Qualifiers  
1..6273  
/organism="Listeria monocytogenes ATCC 19115"  
/db\_xref="taxon:176281"  
BASE COUNT 2064 a 1047 c 1296 g 1855 t 11 others  
ORIGIN

Alignment Scores:  
Pred. NO.: 2.5e-15 Length: 6273  
Score: 323.00 Matches: 124  
Percent Similarity: 37.93% Conservative: 74  
Best Local Similarity: 23.75% Mismatches: 214  
Query Match: 11.91% Indels: 110  
DB: 6 Gaps: 17

US-10-066-551-4 (1-525) x AX416851 (1-6273)

Qy 4 AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23  
Db 555 AATCCAGACAGACAGAAATTC-----AGCACAGCTCAAGATATGAAAGAACTAAATTA 608  
Qy 24 IleSerLeuLeuCysLeuLeuLeuCysSerCysSerSerTyrProLeuGluGlu 43  
Db 609 ATTAAATGCAATTCACGCAATTCACAAACACATATGAAATTTACCGACGATCTGTAAA 668  
Qy 44 ArgThrGlnSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63  
Db 669 AGAATT-----GCCCATTTTAACGTCATAGAGCCAAATTAAG--GGCAATTAATTCGAA 719  
Qy 64 IleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnProHis 83  
Db 720 ATATTA-----ACAAATGCG-----GAA 737  
Qy 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGln 103  
Db 738 GAAACTTTTCCAGTCTCTTACAGCGCTTAGAAAGGCTGAAACACATTCAATTTCAA 797  
Qy 104 TyrTyrIleTyrPargAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAla 123  
Db 798 TACTATATTTTCAACAGCATGAAATTTCTACAAAGATCGTGATATTTTGTAGAAAAA 857  
Qy 124 AlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArgGlyLeuAsp 143  
Db 858 GCAAAAGCAGCGCTGGAAGTTAGATTATGTTGATGCGCTTGATCAAGCAAACTGGGT 917  
Qy 144 AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProHe 163  
Db 918 AAAGCTTTTATGACCTCTTGAAAGAGACCTGAGATGATATTCACGCAATTTGACCA-- 974  
Qy 164 ValLeuArgLysTyrPargAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183  
Db 975 ATTACTCTCTCCGTGGATTTGA-----AGAACGCCAATTTTAA 1013  
Qy 184 MethisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIle 203  
Db 1014 AATCACCGTAAATTTGCTGATGATGTCGCAATTTGCTTACAGGCGGCTTAATATTT 1073  
Qy 204 GlyAspGluTyr-----PheLysValGlyLysAspThrValPheAla 217  
Db 1074 GGAGAGATATTCGTTCAATACACACAGATTTCCGCTGTTGGCGGATACA----- 1124  
Qy 218 AspLeuAspIleLeuAlaThrGlySerValAlaGlyGluValSerHisAspPheAspArg 237  
Db 1125 -----CATATTAAATGACAGGCCAGCTGTTATGCAACTTCAAGATCTTTCTAAT 1178  
Qy 238 TyrTrpAla-----SerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsn 255  
Db 1179 GATTGGGCTATATGCAAAACCAACAGCTGCTGAGATCAATTTATGAGAGCTGGA 1238  
Qy 256 IleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspArgIleThrSerArgHisAlaLeuLeu 275  
Db 1239 TTACACAATATTTTTCACCACTGATGCTGGGTGACGAATGGCG----- 1283  
Qy 276 ArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIle 295  
Db 1284 -----CAAGTGATTTTACGGGAGCCATAT 1307  
Qy 296 AspTrpGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315  
Db 1308 GAT----- 1310  
Qy 316 AspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLys 335  
Db 1311 -----AAGAAAGAAAGGCTCGCGATTCGATTCGTTGATTAAATTCGCTTAAGAA 1364  
Qy 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355  
Db 1365 TCTGTTGGATTGTGTCCTTACTTGTTCAGATGAGAGATCCCTTCGATTCGCTTATTCG 1424

OY	356	LysleuValAlnspgIyeAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp	375
Db	1425	CGGGTCCCATGCATGCCGTGGTCAGCGATTATATTCACAGC-----AAAGCTGAT	1478
OY	376	ValAlaAlaValHisSerGlyTyrrValLysTyrrArgLysProLeuLeuLysAlaGlylle	395
Db	1479	CGCGGATTCATTCACACGGAAGTAATGCGTATAAACAATGATTGACACGCGCG	1538
OY	396	LysLeuTyrrcLuleuGlnProAsnHISAlaValProAlaThrLysAspLysGlyLeuthr	415
Db	1539	AAAATGATCTTATGCCGATGCTTTGTT-----	1571
OY	416	GlySerSerValThrSerLeuHISAlaLysThrPheIleValaspGlyLysArrglePhe	435
Db	1572	-----CATGGCAAGGCATCTAGTGACGGACGGACGGCTCACGCT	1610
OY	436	IleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetcLylVal	455
Db	1611	ATAGGATCTCCCATTTTGCCTGCTAGTTTAGATGATCAATCAATTAATGATATTC	1670
OY	456	IleGlySerProLysIleaglIngluMetgluArgThrLeuAlaAspThrSerProgu	475
Db	1671	TTCATGATGATA-----ACGAGGCTATGTCATCATTTAAACGTCATTTTGAGAAAGAT	1724
OY	476	TyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuglnTrpHisAspProAlaThr	495
Db	1725	TTT-----GAAGATAGCCGACT-----TTTACG	1748
OY	496	ArgLysThrTyrrProAsnglUrogluAlalLysLeuTrpLysArgIleAlaLysIle	515
Db	1749	ATGAAAGATATGGAAACAACCATTTATGACTCGTATAAAGAAGTTCTATCTAGATTTA	1808
OY	516	Leuser 517	
Db	1809	CTATCA 1814	
RESULT 34			
LOCUS	AL596163	231450 bp DNA linear BCT 06-JUN-2002	
DEFINITION	Listeria innocua ClpI1262 complete genome, segment 1/12.		
ACCESSION	AL596163		
VERSION	AL596163.1	GI:16412421	
KEYWORDS			
SOURCE	Listeria innocua.		
ORGANISM	Listeria innocua		
REFERENCE	Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.		
AUTHORS	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A., Baquero P., Berche P., Bloeker H., Brandt P., Chakraborty T., Charif A., Cheouani F., Couve E., de Darvar A., Denoux P., Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussanquet O., Entian K.D., Eschl R., Portillo F.G., Garrido P., Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D., Jones L.M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurakat A., Madueno E., Maitournam A., Vicente J.M., Ng E., Norclari H., Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.C., Purcell R., Remmel B., Rose M., Schlieter T., Simoes N., Trieretz A., Vazquez-Boland J.A., Voss H., Wehlund J. and Cossart P.		
TITLE	Comparative genomics of Listeria species		
JOURNAL MEDLINE	Science 294 (5543), 849-852 (2001)		
PUBMED	21537279		
REFERENCE	2 (bases 1 to 231450)		
AUTHORS	Glaser P., Frangeul L. and Rusniok C.		
TITLE	Direct Submission		
JOURNAL	Submitted (09-JUL-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE		
COMMENT	E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86. Location/Qualifiers		
FEATURES	1..231450		
source	/organism="Listeria innocua"		

[illegible]

gene	transl_table=11	/protein_id="CAC95241.1"	/db_xref="GI:16412429"	/db_xref="SPTREMBL:Q92F05"	/translation="MRKLOPELFYAVVLFVEYVLINQAAILEFLISGIILOLGVIRLLFDQNRNTSKKAAVAVIFILPVLGTTISVLFQGNRPTRKSTQAYWEKAKIHAIPIPNTEKLPRLSKRIIAH/TSLEIPKJKEILINGEFEFPVLLDLRAKEEIOYVYKTDIAISTEIRDLIVLEKAKSGVAFVRPMDDGSSLSKAPLAKKGVYGFPIASPTVAPRANRNRKIVVIDGOIGFEGGLNGEESRSTNPPRWRNTHGQAVTELOESFLNDITVYENQAGADQIFSAIGIQQTFSPVDGDMAYVITGGKEKWRSDMSKMLIDSAKESVWIVSPFVPEDEBALAVIRKAASSVDVRLVTPKGISEFSGSAAYVKTALIEAGAAVYVADSPVHAKMALVQGTALAIQTAIVDRSFRLLMPEYLYESAMLIHLKRFDKPDESRFLTFMKDMENPFLIRIEVLSLSPLIL12247.1.12777
RBS	/gene="11n0009"	12247.1252	/gene="11n0009"	12259.12777	
CDS	/gene="11n0009"				
Alignment Scores:					
Pred. No.:	6.58e-13	Length:	231450		
Score:	318.00	Matches:	123		
Percent Similarity:	38.128	Conservative:	76		
Best Local Similarity:	23.568	Mismatches:	213		
Query Match:	11.728	Indels:	110		
DB:	1	Gaps:	17		
US-10-066-551-4 (1-525) x AL596163 (1-231450)					
QY 4	AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23				
Db 10972	AATCCACCACTGAAAAATTC-----ACGACGCTCAAGTGAAGAAAAAGCTAAATTA 11025				
QY 24	IleSerLeuLeuLysLeuLeuLysSerCysSerSerTrpLeuProPheGluGlu 43				
Db 11026	ATTATATCGCATTCATGCATTCACAAACACACCATGAAATAATTACCAAGCTTTCTAA 11085				
QY 44	ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63				
Db 11086	AGAATT-----GCGCATTTAAAGCTTCATGACGACGCGATTAAAG--GGAAATTAATAATTGA 11136				
QY 64	IleArgHisThrProHisThrAsnGlnLysLeuSerAspIleTyrLeuLeuAsnProHis 83				
Db 11137	ATATTA-----ACAAATGCT-----GAG 11155				
QY 84	GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGln 103				
Db 11155	GAACACTTTTCCAGCTGCTCTACAGCGGCTTCAAAAAGCTGAAACCAATTCACATTCA 11214				
QY 104	TyrTyrIleThrPArgAsnAspIleSerGluArgLeuLeuPheAsnLeuValTyrLeuAla 123				
Db 11215	TATTACATTTTTAAACGCGATTCGATTTCACAGAGATTTCGATATTTTGGTGGAAAA 11274				
QY 124	AlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArgGlyLeuAsp 143				
Db 11275	GCGAATACCCGCTGTAGGTTAGCTTANGTTAGACGCTTGATGACAAACCTGAGT 11334				
QY 144	AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163				
Db 11335	AAAGCTTTTACTCTCTTGAAGAAGAGCTGAGTAGTATTCACGATTTGAGCCA--- 11391				
QY 164	ValLeuArgTyrTrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183				
Db 11392	ATTGCTTCCTCCCTGGATTGA-----ACAAACGCCCAACTTAAGA 11430				
QY 184	MetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlnGlyArgAsnIle 203				
Db 11431	AATCACCGTAAATGTGTGATGATGACGCAAAATTTGCTTACAGAGTGACCTTAACATT 11490				
QY 204	GlyAspGluTyr-----PheLysValIleGlyLysAspThrValPheAla 217				
Db 11491	GGGGAAATATATGCTCTATACACCAAGATTTCGTGTCTGCGCTGATACA----- 11541				





[illegible][illegible]





OY	316	AsparGAGLySProPDIleAglrYLgLeuGlInASpAlaleUySGlnPrOGLuYS	335
D607	-----AAAGAAAATGGGTTCGGATTCCGATTCGATTTTGATTCATTCGCCCTTAACA	6120	
OY	336	SerValTYrLeuValSer-ProTYrrPhelValProThrLYSSergLYThrASpAlaleUla	355
D6121	TCAGTTGGATTGGATGCACCCCATTTTCCTCGCAGATGAAGAAGTCCCTGGCTTATTCGC	6180	
OY	356	LysLeuValGIASPGLIleAspAlalrVallauTheuthrAnserSeteUGlInAlethrsp	375
D6181	CGGGTGGCAATGACTGCTGGTGCCTGGCGGTATTAATTCCTGGT-----AAAGGCAT	6234	
OY	376	ValAlaAlaValAlHisSerGlyTYrValLYSTyrArqLysProLeuLeuLysAlagLYle	395
D6235	CGCGGACTTCATTCACCAGGAAGTAGAACCTTAGTAATAAACATGATTTGAAGCACGCCG	6294	
OY	396	LysLeuTYrGLueGlInProASnHISAlaValProAlalThrLysASPLsgLYleUTHr	415
D6295	AAAAAGTATGCTTAATGCCGATGACACTTTGTT-----	6327	
OY	416	GlySerSerValThrSerLeuNhisALasTrhrPhelValIASPGILysarqLiephe	435
D6328	-----CATGCCAAGAAGATGCTTGCGAGGACAGACAGCTGGCT	6366	
OY	436	IleGlySerPhenAsnLeuAsPproArgSerAlarqLeuasnthrGluMetelyValVal	455
D6367	ATTGCAACTGCCAATTTTAGCGTGGCGTATTTAATGATTAATCAATGAATTAAGTATTC	6426	
OY	456	IleGlySerPROLYSIIleAlaGluInMeTIUAqrTHrLeuAlaAsPThrSerProGU	475
D6427	TTATATGATGA-----ACGAGGCTTATGCATCTTTAAACGTCGACTTCAAANAAGAT	6480	
OY	476	TYralATYrArqValThrlEuAsprghISanArqLeugInTrPHIsAsPProLaThr	495
D6481	TTT-----CAAGATAGCCAGCTA-----TTTAGC	6504	
OY	496	ArgLYSThrTYRProASnGluProGluAlaLysLeuThryLSarqLIleAlaLyslle	515
D6505	ATGAAAAGATATGAAAACAAACATTAATGTGACTGTATAAAGAAGTTCATACAGTTTA	6564	
OY	516	Leuser 517	
D6565	CTATCA 6570		
RESULT 38			
LOCUS	ALS91973		
DEFINITION	200050 bp DNA linear BCT 06-JUN-2002		
ACCESSION	Listeria monocytogenes strain EGD, complete genome, segment 1/12.		
VERSION	ALS91973 ALS91874		
KEYWORDS	ALS91973.1 GI:16409359		
SOURCE			
ORGANISM	Listeria monocytogenes. Listeria-monoctyogenes Bacteria; Filicutes; Bacillales; Listeriaceae; Listeria.		
REFERENCE	1		
AUTHORS	Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A., Baguero,F., Berche,P., Bloecher,H., Brandt,P., Chakraborty,T., Charbit,A., Checourant,F., Couve,E., de Darival,A., Denoux,P., Domane,E., Dominguez-Bernal,G., Duchaud,E., Durant,L., Dussaugel,O., Entlan,K.D., Fshih,I.H., Portillo,F.G., Garrido,P., Jackson,D., Goebel,M., Gomez-Lopez,N., Hahn,T., Hauf,J., Jacksom,D., Jones,L.M., Keerst,U., Krett,J., Kuhm,M., Kunst,F., Kunaprat,G., Madueno,E., Maitournam,A., Vicente,J.M., Ng,E.F., Medhat,H., Nordisiek,G., Novella,S., de Pablo,B., Perez-Diaz,J.C., Purcell,R., Remmel,B., Rose,M., Schuetter,T., Simoes,N., Tierrez,A., Vazquez-Boland,J.A., Voss,H., Wehlund,J. and Cossart,P.		
TITLE	Comparative genomics of Listeria species		
JOURNAL	Science 294 (5543), 849-852 (2001)		
MEDLINE	21537279		
PUBMED	11679669		
REFERENCE	2 (bases 1 to 200050)		
AUTHORS	Glaser,P., Frangeul,L. and Rusniok,C.		

COMMENT	FEATURES
<p>TITLE Direct Submission</p> <p>JOURNAL Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE</p> <p>E-mail: pglaser@pasteur.fr</p> <p>Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.</p>	<p>Location/Qualifiers</p> <p>1. 200050</p>
source	<p>/organism="Listeria monocytogenes"</p> <p>/strain="EGD-e"</p> <p>/db_xref="taxon:1639"</p>
gene	<p>305..1673</p> <p>/gene="dnaa"</p>
RBS	<p>305..310</p> <p>/gene="dnaa"</p>
CDS	<p>318..1673</p> <p>/gene="dnaa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="Chromosomal replication initiation protein Dnaa"</p> <p>/protein_id="CAC98216.1"</p> <p>/db_xref="GI:16409360"</p> <p>/db_xref="SPTREMBL:O8YAW2"</p> <p>/translation="MSIEDIMQETLQIVKNNKSPSYDTMKSTASHLEGNTFIISAPNNVDMLEKSYTORIANILOETRLDVRIDGDEGNEPFTYIKPMPALDEGGIEIKHMLNRYVFDTRVIGSSGRFPAASLAVAAAKAYNPFTYIGVGLGCTHLMHAHGHIYQOHKDNKAKVMTLSSEKTFEISIRDNKTEERTKTYRNDVLLIDIOFLAEGEGTQEEFFHTNLYDEOKIILISSDRPKIPILEDRLRSFPWGLTIDTPDLEGRILIRKRAKADGLIDINEVWLVIANOIDSNIREDGLIRVAVSSIVNNDITAGLAELAKLIDIPSSKSOVITISGIOEAVGEFVRLRDFAKKRTKSIAPROAMYLISRELTPLASLPKIDGFEGRHTTYIHAHEKISQLKTDQVLYKNDLAIETENKLRQAQNMVF"</p>
gene	<p>1856..3062</p> <p>/gene="dnaa"</p>
RBS	<p>1856..1860</p> <p>/gene="dnaa"</p>
CDS	<p>1867..3012</p> <p>/gene="dnaa"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/product="DNA polymerase III, beta chain"</p> <p>/protein_id="CAC98217.1"</p> <p>/db_xref="GI:16409361"</p> <p>/db_xref="SPTREMBL:O8YAW1"</p> <p>/translation="MKFVIERDRLVQAVNEVTRAI SAR TPIITGIKIVNDEGVTLTGSDSISIAFPIELPIENDEIVVESFGIVLQSKYFCDIVRLPENNEIEVTSNYQTNISQGSQFTLNGLDPMPEYPKLPETYDCKTIKIPVNLKNIYRGVFAVSALEVRPVLGVNNVNIKENKLSAVATDSHRLALKEIPLETIDEEVNIYVIGKSLSEKTLDDAASEIEMTLANNOLLFKLKLILFYRSLGSGSPDNRSLIPDTSKELYINSGAPLOAIDIRASILARENNNVYIKMLTLENGOVYESNSPEGVGSENVFSQSFTEEKISFNKGYMMALRAFEGDDDIQISFSGTMRPEVLRPKAANPEILLQITPVRY"</p>
terminator	<p>3030..3062</p> <p>/gene="dnaa"</p>
gene	<p>3104..4464</p> <p>/gene="lmo0003"</p>
RBS	<p>3104..3109</p> <p>/gene="lmo0003"</p>
CDS	<p>3121..4464</p> <p>/gene="lmo0003"</p> <p>/note="conserved hypothetical protein"</p> <p>/codon_start=1</p> <p>/transl_table=11</p> <p>/protein_id="CAC98218.1"</p> <p>/db_xref="GI:16409362"</p> <p>/db_xref="SPTREMBL:O8YAW0"</p> <p>/translation="MKMDTGTNPTKILIFAMPMLIGNLFOQFTYIMDAVYCKEYSVDLAVGATNSVNFEMISLILGLMSGISVVAQYREFKQYDRKDVIAVATYAVFSAITILAVAYLAKPLILRTPNILIDDSYIFLTLEFIIGLPMSLGNAAIIRALGNSTIPLFLLISLMTALDELFEVYVMDMGVGAIAIVLSOTAAIAIVYIYAAHVPEMRTERPKSLPTSLIKEMVYIGLPSGLGSGFISGNMALOSILNGRESSVYAAVTASRIDSLSYQGIARCAASSMRAGONITAGKIDRAREGWSKIKYVTAISIGITTLVOLFARVSLVYFDSETEVINIGVSTLLVLYFVAVVGLIFVRETLRGDGMVAPLAMIPELVSLHVLGEVISTLIGYVGLMWAVPVAITVYVGLVGLWVSGAMOKKAVIRK"</p>

```

gene      join(4631..4637,4644..4865)
RBS       /gene="lmo0004"
          4631..4637
          /gene="lmo0004"
          4644..4865
          /gene="lmo0004"
          /note="similar to B. subtilis Yaag protein"
          /codon_start=1
          /transl_table=11
          /protein_id="CAC98219.1"
          /db_xref="GI:16409363"
          /db_xref="SPTREMBL:O8YAV9"
          /translation="MAETVKINSEFTVLGOLLMDIVYSTGMAKATLSENTIYINCE
ODNRKKLRNGDYLIVGVGKATIEGK"
gene      4856..4861
RBS       /gene="RecF"
          4856..4861
          /gene="RecF"
          4869..5981
          /gene="RecF"
          4869..5981
          /gene="RecF"
          /codon_start=1
          /transl_table=11
          /product="RecF protein"
          /protein_id="CAC98220.1"
          /db_xref="GI:16409364"
          /db_xref="SWISS-PROT:O8YAV8"
          /translation="MHLESLIVLRNFNVNMLEFSPVNVFLGENAQGTNLEAVL
MLAKSHRTNDKDFIMMEKEAKMGRIAKHOSVPLELATIQRKRAKVNLEQK
KLQYVGNLVNVIAPEDLSLVGAPGIRRRFLNMFGLMOPITLHLSRYORLQOR
NOYLKMQKRKVDPIILDLITRQFNAVAINLTRRADFTOKLEAYAPRHHQSRSL
ETIKETKASITLNGDDPEYWKADLLQKMSIKOREIDRGVTLIGPRHDSLEFINQ
NVDFSGQGGQRTTALSIKLAIEDLHEETGEFVLLLDVLSLDDYRQSHLLGALF
GKYQTFVTSTGIDHETLKQATTFYVERGYKKS"
gene      6017..7970
RBS       /gene="gyrB"
          6017..6021
          /gene="gyrB"
          6030..7970
          /gene="gyrB"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase subunit B"
          /protein_id="CAC98221.1"
          /db_xref="GI:16409365"
          /db_xref="SPTREMBL:O8YAV7"
          /translation="MSEENITVQENASDYNEDQIQVLEGLAEVRRKPRMYIGTSOR
GLHLWEIYDNAIDEALAGCTEITITADNSITVRNGRGIPTGINKIGRPVVE
VITFVLHAGKPGCGGKYGSGLHGVASVYNALSTSLRYVHREGKYYQRFREGYV
VMDMEGGETDYRGTIVHTPPDPIETFEFDVLTIRIRLAFINRGLTISTEK
REHKKVRKDHYEGGIRSYVEHLNKAQDVHEPITYLEGERDNLISMQYNTGFS
NIISFANNIHTYEGGTHESGFKTALTRVINDYARNKLFKDSDDLGEDEVREGTAT
ISIKHPDPEQGTQKTKLGNSEARSITDKLFSEALNKEMFNPPVAKIYEKGVAVAR
ARLAARKARREVARSGSLSEISLPGKADCSRNPEISELYIVEGDSAGSAGGRDR
LFOALIPRKLINVERAKLRDRIIANEIRITIPFAMTSGFGDPDYSKSKHIIIMT
DAYVDGAHITLITLLETFRYMRPLDGGITYIAQPRYLQIKHQIETIVSDGLEHY
LASLDGTFKTSIQRYKGLGEMNEPEQLMDTTPNBEHRTLLOVINKDAIDADTPEFMNG
DVEPRRRKTEDNAQYVKNLDY"
gene      8048..10653
RBS       /gene="gyrA"
          8048..8053
          /gene="gyrA"
          8065..10593
          /gene="gyrA"
          /codon_start=1
          /transl_table=11
          /product="DNA gyrase subunit A"
          /protein_id="CAC98222.1"
          /db_xref="GI:16409366"
          /db_xref="SPTREMBL:O8YAV6"
          /translation="MAETPNQRIEINLNKEMRPTSFUDYAMSVIVARALPDVRDLKP
VHRIIYAMNDLQWTSKAYKKSARIYGEVIGKYPHGDIAVYFTWVRMAQDPFYRNK

```

```

LVDSHNGSVSDGMAAAMRYTEARMKISKELLRIDNKDTIDYADYDGSEREPIVL
PARRPNLLVNGSSGIAVGMATNIPHLHGEIVDVALSHDPEITIDIMLEYJGPPE
PTAPMIGRSGSIRRAVESGRSITVRNDEEKNKNETIETIIEIOYVAKALVR
IAELARKKIDIGTISLNDSDSRGMRIVIEVRIRISASVYNNFKKATLOTPTGIMA
LAIVDNHPRKILNKEILLIYYLHEKHOKVIRRRTEELKARARAHILEGRLALDNIDA
IKILRSKTSIDVAKREGLMQFNISDKOAOAILDMQRORLTGLEREITEEYQVVAL
INDKALADDERLLEITIRELEIKKIVAKRRTELLAGDVLSEDEDLIPEEVAI
TLTKRGYIKRLPLSTYRSQRGRGIGQMSHEDFVEHLVAVSTHDTLLFTNTGKY
YRSKGYEPEYGRYAKGIPILNLIGIESQEQVNAVIMLSEPTDSYLFETTKHGVVR
TTLSQFAKIRSGSLRAVELRENDLISVQMTDGSKNMIATKHQSIFYPEENIRVWG
RTAGVAGRIRLREDEVIYGEVLEDEKRVIVTEKYGKQKOPASQYPLRNKGWVKT
VTIREKNGINLYAMKTVTGVEDIMMTYSGVILIREIDTVSGTGSAGVNLIRIENDE
KATVAVAKPREDEVELEBEETIDTLITIQVDESEDPDPSDIEE"
terminator
10600..10653
gene      /gene="gyrA"
          10712..12242
          /gene="lmo0008"
          10712..10717
          /gene="lmo0008"
          10728..12242
          /gene="lmo0008"
          /note="similar to cardiolipin synthase"
          /codon_start=1
          /transl_table=11
          /protein_id="CAC98223.1"
          /db_xref="GI:16409367"
          /db_xref="SPTREMBL:O8YAV5"
          /translation="MKRIQIFLFAVVLFLVEYVLINQAAILFLVTSGIQLGCVIIA

Alignment Scores:
Pred. No.:      1,92e-12      length:      200050
Score:          311.00      Matches:      123
Percent Similarity: 37.36%      Conservative: 72
Best Local Similarity: 23.56%      Mismatches: 217
Query Match:     11.46%      Indels:      110
DB:              1          Gaps:      17

US-10-066-551-4 (1-525) x AL591973 (1-200050)

OY      4      AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      10971  AATCAGCAGCACTAGCAAAATTC-----AGCACAGCTCAAGTAAGGAAACAAATTA 11024

OY      24     IleSerLeuLeuCysLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 43
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11025  ATTAAATGCGATTCATGCAATTCGCAAAACACTAATGAAATAATTCGCCAGCTATCAAAA 11084

OY      44     ArgThrGlnSerArgGlnIlePheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11085  AGAATA-----GCCCATTTAAAGTCCTATAGACCAATCAAG---GGGAATATAAATGCAA 11135

OY      64     IleArgHisThrProHisIleThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHis 83
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11136  ATATTA-----ACAAATGCT-----GAG 11153

OY      84     GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGln 103
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11154  GAAACTTTTCAGTGGCTCCTTAGACGCGGCTGTGAAAGGCTGAAACACCATTTACATTGAA 11213

OY      104     TyrTyrIleLeuPargAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAla 123
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11214  TATTTCATTTTCMAAGCGATTCGATTTCTACTGAAATTCGATTCGATTTTGGTGGAAAG 11273

OY      124     AlaGluArgGlyValArgValArgLeuLeuAspAspAsnAsnThrArgGlyLeuAsp 143
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11274  GCGAATCTGAGTGTGAGGTTAGATTATGTTTGATGTCTCGGATCAAGCAAACTGGCGT 11333

OY      144     AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProHe 163
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      11334  AAAGCTTTTTCATCCCTTTGAAAGAGGCTGAGATTAATTCACGATTTTGATCCA--- 11390

OY      164     ValLeuArgLysTrpPargAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183
          |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

[illegible]

```

/gene="XAC0013"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/transl_table=1
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAM34905.1"
/db_xref="GI:21106052"
/translacion="MPTWNRAGCATIPLINGELVHMFLELMALGALVIAESTPHW
LVAVSLVALLLLMTWRGFWPASHRMGDTORDTSSMIDPAELRLRELAEKSAALAA
ARDEPDA"
complement(3893). .5353)
/gene="c1s"
/note="XAC0014"
complement(3893). .5353)
/gene="c1s"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/genemark"
/transl_table=1
/product="cardiolipin synthetase"
/protein_id="AAM34906.1"
/db_xref="GI:21106053"
/translacion="MLAAGAWEMLATIPHEALLIAVYLLVIMAGIVLQKREP
VATLSVLSLAALPEYIGLLIYWGDPKVRQRIRRRSRSGMESVSYCPDPADTGE
LAKIAOSTGTGAPSSATSEVHMLVDGATVAALIEALIGADHJLHEVYIDPQHSGLA
ICALMERARAGVYRLRLMDAIGSSAMTRALRLREAGVETAMFHFHSQILKPFKPM
LNLTHRKRVIVIDGRIGFTGGINVTTDENQVRQAVRDJLHRLQGHVVSJLQVLE
DMVATISQGRNAFPHGOQLMPDDMTFAGQGVDAOVLYSPGSSMEA1HRLVMAIHEA
KHRWLVTPYFVEPGEARMA1LSAALGGLVRLVLYPVSDSLVTAARSPDELLA
GVRIYEGPMLHTRKALLADDVDCIVSANFDSRSPRLNLSLFRDOAVAAEMAGL
IGTDQQAQDEVQPARHRPLWRSRLPEAFARLSPL"
5796. .6455
/gene="XAC0015"
5796. .6455
/gene="XAC0015"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM34907.1"
/db_xref="GI:21106054"
/translacion="MAPPGIOEOPVSQADADQFAIYSVYSYGINIGTLNVK IDLNG
TVPASISRMVEASDEASAFSAFDVRREGARROOVLQLEKNDQVPTNGPFIPE
IOCVLSNAYKMRSEDEYSIRIDTFPNKSAIRAYASAKAIYYVAAYASAGULITS
SSEIYNVYSIESPFAASRPDYQRIQIDWLTSGAATTSGIVTVSGTVLGVDDVQV
K"
6455. .7045
/gene="XAC0016"
6455. .7045
/gene="XAC0016"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM34908.1"
/db_xref="GI:21106055"

gene
6455. .7045
/gene="XAC0016"
6455. .7045
/gene="XAC0016"
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAM34908.1"
/db_xref="GI:21106055"

Alignment Scores:
Pred. NO.: 7.07e-14 Length: 11017
Score: 308.50 Matches: 132
Percent Similarity: 38.22% Conservative: 74
Best Local Similarity: 24.49% Mismatches: 238
Query Match: 11.37% Indels: 96
DB: 1 Gaps: 16

US-10-066-551-4 (1-525) x AE011624 (1-11017)
Oy 11 ProSerGIuThrIleSerLeuMetIcyThrArgSerLeuIleSerLeuLeuLeu 30
||| ||||| ||| ::| ::| ::| ::| |||||
Db 5224 CCGGTCGACCGACGCTGAGCTGGGTGTTCGCTGGCGGCGCTCCGTACTGAGGCTGCTG 5165
Oy 31 LeuCySSerCySerSerTrrIleuProProIleuLu----- 42

```



```

Db 5164 ATC-----TACTACTGCGCTCGCGCGCAAAAGTCAAAAGCGAGCGGCTGCGGCC 5114
Oy 43 -----GluArgThrGluSerArgHisPheAsnThrSerLysProValArgLysAspAsn 60
Db 5113 GGGCGCTCGCGCTCGCGCGTGAAGACGTACAGAGCGTGTGCCCGCGGATGCCGATTCG 5054
Oy 61 IleLeuGlnIleArgHisThrProHisThrAsnGlyLeu-----SerAsp 75
Db 5053 ACCGAGCTGGCCAGATGCCAGTCCACACCGCGGTGCGCCCGCAGCGCGCCGACCGAA 4994
Oy 76 IleTrpLeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSer 95
Db 4993 GTGCACTGCTGTGGTGAAGCGCTGCCACGTACGCGCGCATGAGAACCGATCCGGGCG 4934
Oy 96 AlaGluHisSerLeuAspLeuGlnIleTrpTrpIleTrpArgAsnAspIleSerGlyArgLeu 115
Db 4933 GCACGCGACCATATCCATCTTGGAGTACTACTTCCACCGCCGACACACGCGCGCGCGG 4874
Oy 116 LeuPheAsnLeuValTrpLeuAlaAlaGluArgGlyValArgValArgLeuLeuAsp 135
Db 4873 ATCTCGCGTGGCTGATGGAACGTGCCCGTGGCGGGGTCAAGGTGGCTGTGATGAT 4814
Oy 136 AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155
Db 4813 GCGATCGGCTCATCGCGCAATGACCGCGCGCTTGCGCGACGCGCGAGCGCGGGGTG 4754
Oy 156 GluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTrpLeuThr 175
Db 4753 GAACCGCGCTGGTCCACCCCGCGCATGTG-----CTCAAG 4718
Oy 176 Asphe-----ProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsn 193
Db 4717 CCATTCAGCGCGCGCTGTGTGAACCTGCGCACCGACCGAAGTATGCTCATCGATGCG 4658
Oy 194 ArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlyTrpPheLysValGlyLysAsp 213
Db 4657 CGCATCGGTTTCACCGCGTGCATCAACGTACCGACGAGACGAGACGAGGTGCGCAAG 4598
Oy 214 ThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyLysValSerHis 233
Db 4597 CAGCGGTATCGGATCTGCACGTGCGCCGCAAGGCCATGTATGTCGCGACGTGCAAGT 4538
Oy 234 AspheAspArgTrpTrpAlaSerHisSerAlaHisAsnAlaThrArgIleLeuArgSer 253
Db 4537 GTGTTCGAGAGATGTGGCT-----TATGCCACACCGCAGGCGCGCGCA 4493
Oy 254 GlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTrpAsnAspGlyThrSerArgHisAla 273
Db 4492 GCTTTCACGCGCCACAGCTA-----TGCGCCGACGATATGCCACCGCGCGG 4445
Oy 274 LeuLeuArgTrpArgGluThrValGlnGlnSerProLeuTrpGlnLysIleGlnThrGly 293
Db 4444 -----CAGGCGACGCTCATGCGCGCAGGTCGT-----GTGCGGCGTCCG 4406
Oy 294 ArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeu 313
Db 4405 GATTCGTGCTGGAGGACGATCACCGCTGATGTCGCG-----4367
Oy 314 AspaArgAspArgArgLysProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnPro 333
Db 4366 -----GCCATCCATGAGGCC 4352
Oy 334 GluLysSerValTrpLeuValSerProTrpPheValProThrLysSerGlyThrAspAla 353
Db 4351 AAGCAGCCGCGTGGCTGTGCTCACTCACTCTGTCGCCGCGGAGCGCGCGTATGCGG 4292
Oy 354 LeuAlaLysLeuValGlnAspGlyLysAspValThrValLeuThrAsnSerLeuGlnAla 373
Db 4291 CTGACCTGCGCGCGCTGCGCGGTGAGCGTGTGTCGCG-----CGGTC 4238
Oy 374 ThrAspValAlaAlaValHisSerGlyTrpValLysTrpArgLysProLeuLeuLysAla 393

```

```

Db 4237 AGCGATTCCGCGCTGTCACCTACGCGGACGCGCTGATTTTCAGACGAGTGTGAAGCC 4178
Oy 394 GlyIleLysLeuTrpGluLeuGlnProAsnHisAlaValProIleThrLysAspLysGly 413
Db 4177 GGTGTCGCATCTACAGATGACGGCGCGCG-----4148
Oy 414 LeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArg 433
Db 4147 -----ATGCTGCACACCAAGCGCGCTGTGCGCGACGACGATGTC 4109
Oy 434 IlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGly 453
Db 4108 TGCATGCTGGCGCGCGCCCAATTGCGACGCGCGAGCTTCGCGCTGATTCGAGCTGTG 4049
Oy 454 ValValIleGluSerProLysIleAlaGlnGluMetGluArgThrLeuAlaAspTrpSer 473
Db 4048 ATGCTGTTCCGCGACCGCGCGGTGGCGCAGAGATG---GCCGGCTGATGCGCACCGAT 3992
Oy 474 ProGluTrpAlaTrpArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspPro 493
Db 3991 CTGCACACGCGCGAGGAGATGCAATTCGCGCGACCGCGCGCTTA---TGGCGCTGCGG 3936
Oy 494 AlaThrArgLysThrTrpTrpAsnGluProGluAlaLysLeu-----507
Db 3935 CCGCCCGCAAGCGTTTGGCGCGCTGTGCGCGCTGTCTGTGACCAACGCTGCGGTGCC 3876
Oy 508 ---TrpLysArgIleAlaAlaLysIleLeuSerLeuLeuProIleGluSerLeuLeu 525
Db 3875 ACCTGGAACCGC---GCCGGCGCGCGCTACCTGCGCGCTCAACGAGGAGCTC 3822

RESULT 40
AP004829 304050 bp DNA linear BCT 02-JUL-2002
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus MW2 DNA, complete genome,
strain: MW2, section 8/10.
AP004829 BAO00033
ACCESSION
VERSION
AP004829.1 GI:21205117
SOURCE
Staphylococcus aureus subsp. aureus MW2 (strain: MW2) DNA.
Staphylococcus aureus subsp. aureus MW2
ORGANISM
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A.,
Nagai, Y., Iwano, N., Asano, K., Naito, T., Kuroda, H., Cui, L.,
Yamamoto, K. and Hiramatsu, K.
Genome and virulence determinants of high virulence
community-acquired MRSA
Lancet 359 (9320), 1819-1827 (2002)
JOURNAL
MEDLINE
22040717
PUBMED
12044378
REFERENCE
2 (bases 1 to 304050)
Aoki, K., Oguchi, A., Nagai, Y., Asano, K., Iwano, N., Baba, T.,
Kuroda, M., Hiramatsu, K. and Kikuchi, H.
Direct Submision
Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center, 2Chome 49-10
Nishihara, Shiba-ku, Tokyo 151-0066, Japan
(E-mail: oguchi@nite.go.jp, URL: http://www.bio.nite.go.jp/,
Tel: 81-3-3481-8423, Fax: 81-3-3481-8424)
FEATURES
source
1. 304050
/organism="Staphylococcus aureus subsp. aureus MW2"
/strain="MW2"
/db_xref="taxon:196620"
complement(629..1072)
/gene="MW1947"
complement(629..1072)
/gene="MW1947"
complement(629..1072)
/note="ORFID: MW1947
hypothetical protein, similar to acetyltransferase
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=11

```

gene  
/protein\_id="BAB95812.1"  
/db\_xref="GI:21205118"  
/translation="MSITIRLFNNSDEKLNQLOCKLYDGLGYPTNENDLKRRLKITT  
HDDYFLLLIKENKIGISGCKMKFEYKNAEYRILAEVIFHSEFRKKGYGKRLIADS  
EEFSKRLNCAITLNSGNRNLSAHLKYSNGVSNVSGFTKOL"  
/complement(1178..1696)  
/gene="MW1948"  
/note="ORFID:MW1948  
[Genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical peroteiin"  
/protein\_id="BAB95813.1"  
/db\_xref="GI:21205119"  
/translation="MDNINKNEKFEVLSTIAITISFILNGLFISLAIMWSNGMKNK  
CISLNLSDPTFAVISMTTIGTALIGTAPISYIGTPTSIDKSSRAISDMDEEO  
RNNKIDMLMLPFLSLIVIFIFQFAYSKLENTAIIIFIIILVVMFLLILICRK  
ITWICRFYSKE"  
/complement(1932..2522)  
/gene="MW1949"  
/complement(1932..2522)  
/gene="MW1949"  
/note="ORFID:MW1949  
,hypothetical protein, similar to bacteriophage terminase  
small subunit [Genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB95814.1"  
/db\_xref="GI:21205120"  
/translation="MSELTAKQARFVNEYITLWVTSATKAGYNANSAHTACRLK  
KPHIKTIOEKDKTIDENVTAKELHLVLTNAIVGDETEKEVVRKGEYKEMPQS  
KVQIVYNEHVELLEVPKIPSDRLKAREMLKYHLFDKHDINGNVPIFLINIGMDGS  
DEGLDKAVKDVNSDNPNTYIVDDIPLEDYESIRFL"  
/complement(2519..2734)  
/gene="MW1950"  
/complement(2519..2734)  
/gene="MW1950"  
/note="ORFID:MW1950  
[Genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=11  
/product="truncated hypothetical protein"  
/protein\_id="BAB95815.1"  
/db\_xref="GI:21205121"  
/translation="MTDNDVNHIIKNHIEIVRLATKPEYIFDSLIOGLSYODI  
SSALDCSEQSVIIWYETILDKIVGVIK"  
/complement(3338..3338)  
/gene="truncated int"  
/note="ORFID:MW1951  
[Genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB95816.1"  
/db\_xref="GI:21205122"  
/translation="MKWSDIDFENNTISITKYYPNNKKKFOILPPTKSSIGKIS  
VDNIVLVLDYKINVOQNMKNNELYNHCKEKTINVDTNHYVYQHTOKLVALLIGI  
HTRLIARICRTH"  
/complement(3372..3372)  
/gene="truncated int"  
/note="ORFID:MW1952  
truncated hypothetical protein, similar to integrase  
[Genomic island nu Sa 4mw]"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB95817.1"  
/db\_xref="GI:21205123"  
/translation="MITVMKKLSMWIQTIMSRPISILKNISTHSFRTHCALLIESGV  
HIKEIQRLRHKIDINTMNIYAKITTSYKDKASHKFSHOMEDVS"

gene  
/complement(3433..5049)  
/gene="groEL"  
/complement(3433..5049)  
/gene="groEL"  
/note="ORFID:MW1953"  
/codon\_start=1  
/transl\_table=11  
/product="groEL protein"  
/protein\_id="BAB95818.1"  
/db\_xref="GI:21205124"  
/translation="MWKOLKRFSEDAQMLRGVQDILANAVKVTIGPKRNVYLKKEFT  
APLTNDGVTITAKLELEDRENNGAKLYQVANKTEINAGDGTITNTVLAQAMIOES  
LKNVTSKANPVGLQGDIDKAVVAEALHENSQVKNKNEIAQVGAISADEETGRIT  
SEAMEKVGNDVITIEESNGLTELVEVGEQFQSGPYMTVDSKMAVALERPPI  
LVTDKISSFQDIDILPLEOVVQSNRPILIVADEEGALNTVINRMKGFETAVAVA  
PGFDDRKAMLEADAILTGAQVITDDGLDKASIMLGTAASKVEYTKONTVYDGD  
GDENSIDARVASQLSQIEETSDPDREKLOERLAKGLAGVAVIKVGAASETELKERL  
RIEALNISTARVAEEGIVAGGCTALVAVYQKVSFTEAGDIEGCVNIVKATLPVVO  
IAENAGLEGSIVTERLKNABEGVFNATNEMVMLEAGIVDPKTVRSALQHAASVA  
AMFLTEAVVASIPEKNDOPNMGMPGM"  
/complement(5125..5409)  
/gene="groES"  
/complement(5125..5409)  
/gene="groES"  
/note="ORFID:MW1954"  
/codon\_start=1  
/transl\_table=11  
/product="GroES protein"  
/protein\_id="BAB95819.1"  
/db\_xref="GI:21205125"  
/translation="MLKPIGNRVITIEKKEBOTKSGIVLTDSAKESNECVIYAVGT  
GRLLNDGTRVPEVEKEDRVYFOIAGIEVAKRDNETVLVNEEDILAVIE"  
/complement(6327..6327)  
/gene="MW1955"  
/complement(6327..6327)  
/gene="MW1955"  
/note="ORFID:MW1955"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAB95820.1"  
/db\_xref="GI:21205126"  
/translation="MTRLMASLITVITIIISQPLILVKKLPPVOYSGIELTKAVIY  
IQVLFLIATITILINKIKNPTRKLEKPEPKYIIPNALISFALMYIOMVSYIY  
LTOIYGGQOVSPNTEKLIILARKPIPTFEVSIIGPLEVEVFKVIFGELFNKIKN  
RIVAFIATVYSSLIFFALAHNDFKPIFYVFGMGVIFSLAVYMKRLAVPIIIMLONG  
FVVFQOLNPEALKATEQANFIYHIFP"  
/complement(6352..7611)  
/gene="MW1956"  
/complement(6352..7611)  
/gene="MW1956"  
/note="ORFID:MW1956  
hypothetical protein, similar to sdh"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB95821.1"  
/db\_xref="GI:21205127"  
/translation="MSYHWEKKMLSTSIILSSSGLATHTVEAKONINGEKPTTN  
LNHNITSPVNSEMMNNNETGPHESNOTGNETGSNSRDANPDSNNVKKPSSNNO  
DSDKPDNNONSPPKPDNDPKPKPPKPPDPKPPKPPDPKPPDPKPPDPKPPD  
DKPKPDNDPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPKPPK  
ASDGSNOGQMPNONGSONPTGNDVYSQFLALANGAKYKRYILNQINKLGGKDG  
EVIDEDYINIIIRKONFSGNALVNLQOOSNFRQYFNPLKSEYRYNLDQVALIT  
GETSGMPDLKPKEDKPDKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPS  
MMVVSIMLEFLFVRNKKKNKKNESQRR"  
/complement(8434..8434)  
/gene="MW1957"  
/complement(8434..8434)  
/gene="MW1957"  
/note="ORFID:MW1957"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"





4946.858 Million cell updates/sec

### Scoring table:

Xgapop	10.0	,	Xgapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
Delop	6.0	,	Delext	7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:	4370478
--	---------

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command Line parameters:

```

-O-cgml2/USFPO.spool-US10066551/rnauc.12052003_091142.22336/app_query.fasta.1.7111
-D-B-n.GeneSeq.101002-USFMO-fastap-SUFFIX-rxn-0-MIMATCH-0-1-LOOCP-0
-LOOPEXT-0-UNITS-bits-START-1-1-END-1-1-MATRX-biosome62-TRANS-human40-0
-LIST-45-DOCLINK-200-THR.SCORE-pct-THR.MAX-10-THR.MIN-0-ALIGN-40
-MODE-LOCAL-OUTFMT-pct-NORM-emit-HEAPSIZE-500-MINLEN-0-MAXLEN-2000000000
-USR-US1006651.@cnc.1.1.200.@rnatc.12052003_091142.22336-NCPU-6-ICPU-3
-NO-ALXPT-NO-MAP-LARGEBERT-NEG.SCORES-0-WAIT-1-ONGLOG-DEV-TIMEOUT-120
-WARN-TIMEOUT-30-THREADS-1-XGAOP-10-XGAEXT-0.5-FGAOP-6-FGAEXT-7
-YGAOP-10-YGAEXT-0.5-DELOP-6-DELEXT-7

```

Database

1:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT *
2:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT *
3:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT *
4:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT *
5:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT *
6:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT *
7:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT *
8:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT *
9:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT *
10:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT *
11:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT *
12:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT *
13:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT *
14:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT *
15:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT *
16:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT *
17:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT *
18:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT *
19:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT *
20:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT *
21:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT *
22:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT *
23:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT *
24:	/SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	2683	98.9	33303	21	AAA81514	N. meningitidis pa
2	2683	98.9	349980	21	AAE21610	Neisseria meningit
3	2629	96.9	1527	21	AAZ54515	Neisseria meningit
4	2599	95.8	1527	21	AAZ54514	Neisseria meningit
5	2584	95.2	1527	21	AAZ54513	Neisseria gonorrhe
6	839.5	30.9	65792	22	AAE28541	Genomic fragment #
7	777	28.6	1551	22	AAE60961	P. putida RT2440-a
8	632	23.3	1509	18	AAV24861	H. pylori secreted
9	632	23.3	1509	18	AAV24861	H. pylori secreted
10	628	23.1	1630	19	AAAX14097	H. pylori GHPO 137
11	323	11.9	2319	24	ABO69094	Listeria monocytog
12	323	11.9	6273	24	ABO71029	Listeria monocytog
13	318	11.7	684707	24	ABO67196	Listeria innocua c
14	318	11.7	3011208	24	ABO69245	Listeria innocua D
15	311	11.5	2944528	24	ABAO3041	Listeria monocytog
16	295	10.9	1488	24	ABN91973	Listeria monocytog
17	293.5	10.8	3011208	24	ABO69245	Listeria innocua D
18	288.5	10.6	1544	24	ABO70540	Listeria monocytog
19	288.5	10.6	2944528	24	ABAO3041	Listeria monocytog
20	280.5	10.3	1446	24	ABK73787	Bacillus lichenifo
21	279.5	10.3	618	18	AAV74658	H. pylori ORF 05cp
22	275	10.1	1482	24	ABAA1653	Staphylococcus aur
23	275	10.1	1509	24	ABAA1659	Staphylococcus aur
24	264.5	9.7	1485	24	ABN91876	Staphylococcus epi
25	264.5	9.7	4403	18	AAV4389	Staphylococcus aur
26	248.5	9.2	28690	20	AAAX13075	Enterococcus faeca
27	243.5	9.0	1389	22	AAAF67781	Corynebacterium g1
28	243.5	9.0	1500	22	AAH67979	C. glutamicum codin
29	243.5	9.0	1850	24	AAAE8810	C. glutamicum ATCC
30	243.5	9.0	349980	22	AAAE8853	C. glutamicum codin
31	242	8.9	18096	23	AAAS9588	Propionibacterium
32	237	8.7	1325	22	AAH52803	S. epidermidis ope
33	237	8.7	2978	22	AAH54735	S. epidermidis gen
34	234	8.7	3745	22	AAH54681	S. epidermidis gen
35	219	8.1	1877	24	ABFL90668	Human polynucleoti
36	208	7.7	640681	24	ABA92787	Buchnera sp. genom
37	204	7.5	1183	18	AAV4787	Staphylococcus aur
38	197.5	7.3	933	24	ABK73802	Bacillus lichenifo
39	176	6.5	3294	23	AAAE5339	DNA encoding novel
40	176	6.5	3294	23	AAAE5338	DNA encoding novel
41	176	6.5	3294	23	AAAS88670	DNA encoding novel
42	176	6.5	3294	23	AAAS88822	DNA encoding novel
43	176	6.5	3294	23	AAAS88993	DNA encoding novel
44	169.5	6.2	1830121	17	AAV42063	Haemophilus influe
45	165	6.1	1575	24	ABN69145	Streptococcus poly

PD 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 XX Frazier CM, Hickey E, Peterson J, Tettein H, Venter JC;  
 PI Masiugnani R, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 PI Rappuoli R, Pizzo M;  
 XX  
 DR WPI: 2000-318079/27.  
 XX  
 PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be  
 PT used in the diagnosis and treatment of *N. meningitidis* infection and  
 PT other *Neisseria* infections, for example, *N.gonorrhoea*.  
 XX  
 PS Claim 7; Page 1375-1385; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic  
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA  
 CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to  
 CC AAA81452 represent *Neisseria meningitidis* MemB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC *Neisseria* bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against *Neisseria* B; against all serotypes;  
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC *Neisseria* B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX Sequence 33303 BP; 7919 A; 8691 C; 8422 G; 8266 T; 5 other;

# Alignment Scores:

Pred. No.: 2,28e-264 Length: 33303  
 Score: 2683.00 Matches: 519  
 Percent Similarity: 99.43% Conservative: 3  
 Best Local Similarity: 98.86% Mismatches: 3  
 Query Match: 98.89% Indels: 0  
 DB: 21 Gaps: 0

US-10-066-551-4 (1-525) x AAA81514 (1-33303)

QY 1 MetArgAlaAsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThr 20  
 DB 1791 ATGGCGCGCAACCCCAAAACAGGCAATGCCGCTGAAACCATATCCGTATGAAAAA 1850  
 QY 21 ArgSerLeuIleSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40  
 DB 1851 CGCGACCTAATTCCTTTATGCCCTCTCTGCTTCATGCTTCATGCTTCACCCCA 1910  
 QY 41 LeuGlnGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsn 60  
 DB 1911 CTGGAAGACGAGCGAGGAAAGCGGTCATTTCAATATCTTCCAAACCGCTCCGCTGACAC 1970  
 QY 61 IleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsn 80

DB 1971 ATCCGCAAAATCCGGCGACACCCCTCATACCAACGGGCTATCCGATATCTATCTGTTGAC 2030  
 QY AspProHisGluAlaPheAlaAlaArgAlaLeuIleGluSerAlaGluHisSerLeu 100  
 DB 2031 GACCCCAACGAAAGCTTTGCCCGCCGCCCTTTATGCAATCTGCCGAACACAGCCTTC 2090  
 QY 101 AspleuGlnTyrTyrIleTyrPArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuVal 120  
 DB 2091 GATTTGCAATACATCATCTGGCGGCAACAGACATTTCCGAGGCTGCTGTCAACCTCGTG 2150  
 QY 121 TyrLeuAlaAlaGluArgGlyValAlaArgAlaArgLeuLeuLeuAspAsnThrArg 140  
 DB 2151 TACCTTGGCGAGACGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2210  
 QY 141 GlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPhe 160  
 DB 2211 GGATTGGACGACCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2270  
 QY 161 AsnProPheValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThrAspPheProArgLeu 180  
 DB 2271 AACCCCTTCGCTTACGAAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2330  
 QY 181 AsnArgGluMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGly 200  
 DB 2331 AACGCGCCATGACACAAATCCCTTTACCGCGCAACCGCGCACCATCATCTCGCGGGA 2390  
 QY 201 ArgAsnIleGlyAspGluTyrPheLysValGlyLysAspThrValPheAlaAspLeuAsp 220  
 DB 2391 CGCAATATCGCGCGAGATCTTCAAACTCGGTGAGACCGATTTTCCGCGGATTTGAGAC 2450  
 QY 221 IleLeuAlaThrGlySerValValGlyGluValSerHisAspPheAspArgTyrTyrPAla 240  
 DB 2451 ATCTCGCCACCGCGAGGCTGTGCGGCAATTCGACGACTTCGACCGCTACTGTGGCA 2510  
 QY 241 SerHisSerAlaHisAsnAlaThrArgIleLeuArgSerGlyAsnIleGlyGlyLeu 260  
 DB 2511 AGCATTCGCGCCCAACAGCCGACCGCATCTCCGACGCGGCAATCGCAAGGCTT 2570  
 QY 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThr 280  
 DB 2571 CAACGACTCGGATCAACAGCAAGAAACGTCAGACCGGCTCTGCTTACCGCGAAAC 2630  
 QY 281 ValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTyrGlnSerVal 300  
 DB 2631 GTCCAGACAGTGCCTCTTACCAAAATACAGACAGATCATCTGCAAGGCTC 2690  
 QY 301 GlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysPro 320  
 DB 2691 CGAACCGGCTCATCAGAGACGACCTGCAAAAGACTCGACCGCGCAACCG 2750  
 QY 321 ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValTyrLeuVal 340  
 DB 2751 CCGATTGCGCGGCGGTGCAAGACGCGCTCAAAACCGCGCAAAACGCTATCTGCTG 2810  
 QY 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAsp 360  
 DB 2811 TCACCTTATTTGTTCCACAAATCCGAGCAACGACGACTGCGAAACCTGTCAGAGAC 2870  
 QY 361 GlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHis 380  
 DB 2871 GGCATAGACGTTACGTTTACCACTCGCTGAGCGAGCGACGATGCGCGCTCAT 2930  
 QY 381 SerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeu 400  
 DB 2931 TCCGCGTATGTCAAAATACGAAACCGCTGCTCAAAACCGGACATCAACCTAGAGACTG 2990  
 QY 401 GlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrLysSerSerValThr 420  
 DB 2991 CAACCCCAACATGCGCTCCCGCACAAAGACAAAGGCGCTGAGCGGACCTCCGTAC 3050  
 QY 421 SerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsn 440  
 DB 3051 AGCTTCACGCGCAAAACCTTCATGTGAGCGGCAACGACATCTTCATCGGCTGCTTAC 3110



Db 274302 AGCATTCGCCGCCACAGACCCAGCGGATCATCCGACGGGACATCGGACAGGCTCTT 274243  
 QY 261 GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuArgTyrArgGluThr 280  
 Db 274242 CAAGCACTCGGATACAGACAGAAACGTCACAGACGGCTCTCGTACCGCAAAAC 274183  
 QY 281 ValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyAlaGlyIleAspTrpGlnSerVal 300  
 Db 274182 GTCCGAACGTGCGCCCTTACCAAAAATACAGACAGATGATGACTGGAGAGGCTC 274123  
 QY 301 GlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysPro 320  
 Db 274122 CGAACCCCTCATCAGAGAGACCCCTCAAAAGAGACTCGACCGGACCGCCCAACCG 274063  
 QY 321 ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuVal 340  
 Db 274062 CCGATTGCGGGGCGCTCAAGACCGCGCTCAAAACGCCCAAAAAGGCTCATCTGCTT 274003  
 QY 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAsp 360  
 Db 274002 TCACCTATTTCGTTCACCAAAATCCGACAGACGACGCGCAAACTGGTCAGAGAC 273943  
 QY 361 GlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaValHis 380  
 Db 273942 GGCATAGACGTTACCGTCTGACCAACTCGCTCGACGGGACCGAGCTTCCGCGCTCAT 273883  
 QY 381 SerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeu 400  
 Db 273882 TCCGCGTATGTCAATATCCGAAACCGCTGCTCAAAAGCGCGCATCAAACTACAGAGCTG 273823  
 QY 401 GlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThr 420  
 Db 273822 CAACCCCAACCTGCGCTCGCCGCCACAAAGACAAAGGCGCTGACGCGGACGCTGTAACC 273763  
 QY 421 SerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsn 440  
 Db 273762 AGCGTCGACGCGCAAACTCTATGTGACGCGCAAAAGCATCTTATGATGTTGTTAAAC 273703  
 QY 441 LeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerProLys 460  
 Db 273702 CTCGACCCCGCTTCGCGCGCTCTCAACACCGAAAGGCGGTTGTTATCGAAAGCCCAAA 273643  
 QY 461 IleAlaGlnGlnMetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgVal 480  
 Db 273642 ATCCGACAGACAGATGAGACGACCTTCCATACACACCGCGCTACCGCTACCGCGCTT 273583  
 QY 481 ThrLeuAspArgHisAsnArgLeuGlnThrHisAspProAlaThrArgLysThrTyrPro 500  
 Db 273582 ACCCTCGACAGGACAAACCGCGCTGCAATGGCAGATCCCGCACCGCAAAAACCTACCGG 273523  
 QY 501 AsnGluProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSerLeuLeuPro 520  
 Db 273522 AACGAACCCGGAAGCCAACTTGGAAAGCATCGCCGCAAAAATCTCTCTGCTGCC 273463  
 QY 521 IleGluSerLeuLeu 525  
 Db 273462 ATAGAAAGTTTATTA 273448  
 RESULT 3  
 AAZ54515  
 ID AAZ54515 standard; DNA; 1527 BP.  
 AC AAZ54515;  
 XX  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria meningitidis ORF 987 partial DNA sequence SEQ ID NO:2977.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.

XX W09957280-A2.  
 FN 11-NOV-1999.  
 PD  
 XX 30-APR-1999; 99W0-US09346.  
 PF  
 XX 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 09-OCT-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masianni V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI: 2000-062150/05.  
 P-PSDB: AAY75753.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 PS  
 PS Claim 7; Page 1396; 1453pp; English.  
 CC AA53015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC have use as antibacterial agents, or antagonists, which may themselves  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 1527 BP; 386 A; 547 C; 325 G; 269 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 9,02e-261 Length: 1527  
 Score: 2629.00 Matches: 508  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 96.90% Indels: 0  
 DB: 21 Gaps: 0  
 US-10-066-551-4 (1-525) x AAZ54515 (1-1527)  
 QY 18 MetLysThrArgSerLeuIleSerLeuLeuGlySerLeuLeuGlySerSerSerTrp 37  
 Db 1 ATGAACACAGCAGCGTATTTCCCTTTATGCCCTTCCTCTGTCATGTTCTCATGG 60  
 QY 38 LeuProProLeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArg 57  
 Db 61 TTGCCCCCCTGGAAGAAAGGACGAAAGCGCTATTTCATTCTTCAATCTTCAAAACCGTCCGC 120  
 QY 58 LeuAspAsnIleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyr 77  
 Db 121 CTGGACACATCTGCAATCTGGACACACCCCTCATATCCAGGGGCTATCCGATATCTAT 180  
 QY 78 LeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlu 97  
 Db 181 CTGTTGAACGACCCCGACGAAGCTTTGCGCGCGCGCGCTTATCGAATTCGCCGA 240



QY 98 HisSerLeuAspLeuGlnTyrTrpArgAsnAspIleSerGlyArgLeuLeuPhe 117  
Db 241 CACAGCCGTCGATTGCATACATCTGGGCAACGACATTTGGCGGACTGTGTTT 300  
QY 118 AsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAspAsn 137  
Db 301 AACCTCGTGTACCTTGGCGGCAAGCGCGTGGCGCTACCGCTGTTGACACCAAC 360  
QY 138 AsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluVal 157  
Db 361 AACACGGCGGATTTGAGAGACCTCTGCTGCGCTCGACACCATCCCAATATGCAAGTG 420  
QY 158 ArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAspPhe 177  
Db 421 CGCGTGTTCACACCCCTGCTGTACGAATAATGGCGGCACTGGGTACTCTGCGACACTTC 480  
QY 178 ProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIle 197  
Db 481 CCGCGCTCAACCGCGCATGCACACAAATCTTTACCGCGGCAACCGCGCCACCATTA 540  
QY 198 LeuGlyArgAsnIleGlyAspGlyLysTrpPheLysValGlyLysAspThrValPheAla 217  
Db 541 CTGGCGGAGCAATATCGCGGACGAATCTTCAAGTGGTGAGACACCGTTTTCGCC 600  
QY 218 AspleuAspIleLeuAlaThrGlySerValValGlyLysValSerHisAspPheAspArg 237  
Db 601 GACCTGGACATCTTCGCGACCGGAGCGTGTGGCGGAATATGCGACAGCTTGCACGCG 660  
QY 238 TyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGly 257  
Db 661 TACTGGCGCAACCATTCGCGCCACAAACCGCACGCGCATCTCGGACGCGCAATCGCG 720  
QY 258 LysGlyLeuGlnAlaLeuGlyTyrAsnAspGlyLysTrpArgHisAlaLeuLeuArgLys 277  
Db 721 AAGGCTCTCAGACGATCGGATACAGCAACAGTCCACACGCGCTCTGCGCTAC 780  
QY 278 ArgGlyLysValGlyLysSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrp 297  
Db 781 CGCGAAACCGTCGAACAGTCCCGCTTACCAAAAATACAGACGAGCATGCTGACG 840  
QY 298 GlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArg 317  
Db 841 CAGAGCGTCCAAACCGCTCATCAGCAGACCTGCAAAAGGACTGACCGCACCGC 900  
QY 318 ArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlyLysSerVal 337  
Db 901 CGCAACCGCGCATTCGCGGCGGCTGCAAGACGCGCTCAACACGCGCAAAAAAGCCTC 960  
QY 338 TyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeu 357  
Db 961 TATCTGGTTCACCTTATTTGTCGCCACAAAATCCGCGCAGACGCACTGGCAAAACTG 1020  
QY 358 ValGlnAspGlyLysAspValThrValLeuThrAsnSerLeuGlnAlaThrAspAlaAla 377  
Db 1021 GTGAGAGCGGATAGAGATACCGTCTGACCAACTGCTACAGCGCAGCGACTTTC 1080  
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeu 397  
Db 1081 GCGCTCATTCGCGCTATGCTAATATACGGAACCGCTGCTCAAAAGCGGATCAACTTC 1140  
QY 398 TyrGlyLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySer 417  
Db 1141 TACGAGCTGCAACCAACCATGCGCTCCGCAACAAAAGAGCGCTGACGCGGAGCG 1200  
QY 418 SerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGly 437  
Db 1201 TCCGTAACACCTGCAAGCCCAAAACCTTCATGTGGAGCGCAACGCACTTTCATCGC 1260  
QY 438 SerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlyLysValValValIleGly 457  
Db 1261 TCATTTCACCTGCAACCGCGCTTCCGACGCGCTCAATACTGAATAAGGCGTGTGTATCGAA 1320  
QY 458 SerProLysIleAlaGlnGlnMetGlnArgThrLeuAlaAspThrSerProGlyLysAla 477

Db 1321 AGCCCAAAATCGCAGAACAGATGAGCGCACCTTGGCGGATCTCAACCGGAATACGCC 1380  
QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497  
Db 1381 TACCGCTTACCTTCGACAGCGCAACCGCTGCAATGACAGATCCGCGCACCGCAAA 1440  
QY 498 ThrTyrProAsnGluProGluAlaLysLeuTrpLysArgIleAlaAlaLysIleLeuSer 517  
Db 1441 ACCTACCGCAACGACCGGCAACCGCAACTTTGGAAACGATGCGCGCAAAATCTCTATCC 1500  
QY 518 LeuLeuProIleGlyLysLeuLeu 525  
Db 1501 CTGCTGCCCATAGAAAGTTTATTA 1524  
RESULT 4  
AAZ54514  
ID AAZ54514 standard; DNA; 1527 BP.  
XX  
AC AAZ54514:  
XX  
DE 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 987 partial DNA sequence SEQ ID NO:2975.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN M09957280-A2.  
XX  
ED 11-NOV-1999.  
XX  
PE 30-APR-1999; 99WO-US09346.  
XX  
PR 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX  
PA (CHIR) CHIRON CORP.  
XX (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI: 2000-062150/05.  
XX P-PSDB: AAY75752.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
XX vaccines and diagnostics  
XX  
PS Claim 7; Page 1394-1395; 1453pp; English.  
XX  
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941  
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent  
CC PCR primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention





Db 1501 CTGCTGCCATCGAAGGTTTATTA 1524  
RESULT 6  
AAF28544/c  
ID AAF28544 standard; DNA: 65792 BP.  
XX  
XX AAF28544:  
XX  
XX 04-APR-2001 (first entry)  
XX  
XX Genomic fragment #31.  
DE  
XX Genomic library: bacteria; human upper airway; otitis media; sinusitis;  
KM bronchopulmonary; endocarditis; meningitis; ss.  
XX  
XX Moraxella catarrhalis.  
OS  
XX  
XX W0200078968-A2.  
XX  
XX 28-DEC-2000.  
PD  
XX  
XX 16-JUN-2000; 2000MO-US16649.  
PF  
XX  
XX 18-JUN-1999; 990US-0140121.  
PR  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Lagace RE, Patterson C, Berg KL;  
PI  
XX  
XX WPL; 2001-041427/05.  
DR  
XX  
XX Genomic library for identifying diagnostic and therapeutic  
PT compositions, and for identifying virulence factors, regulatory  
PT elements and drug targets, comprises Moraxella catarrhalis nucleic  
PT acids -  
XX  
XX  
XX Claim 1; Page 278-293; 545pp; English.  
PS  
XX  
XX The present invention relates to a Moraxella catarrhalis genomic library  
CC comprising of a combination of 41 nucleic acid molecules (see  
CC AAF28514-AAF28544). The library has a number of uses described in the  
CC specification e.g. is useful for identifying diagnostic and therapeutic  
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
CC aerobic, gram-negative diplococcus, normally found among the bacterial  
CC flora of human upper airways. M. catarrhalis is known to cause acute,  
CC localised infections such as otitis media, sinusitis and bronchopulmonary  
CC infection and life-threatening, systemic diseases including endocarditis  
CC and meningitis.  
CC  
XX  
XX Sequence 65792 BP; 18902 A; 12532 C; 14627 G; 19731 T; 0 other:  
SQ  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 5,46e-74 Length: 65792  
Score: 839.50 Matches: 195  
Percent Similarity: 55.08% Conservative: 103  
Best Local Similarity: 36.04% Mismatches: 196  
Query Match: 30.94% Indels: 47  
DB: 22 Gaps: 12  
US-10-066-551-4 (1-525) x AAF28544 (1-65792)  
QY 22 SerleulleSerleulleCysleulleCysSerSerSerleullePro----- 39  
Db 30430 AGTATACATCTGCTGATGCTGTATCTTGTGGTGCAGAAAT--TTGCCGATACG 30374  
QY 40 -----ProleugluuLgrgThrgluserArgHispheasnThr 52  
Db 30373 CCTACTTACCAAGAGTTGGCATGTGACCAAGAGGTTCAAGAGCGTCAACAAATGGTC 30314  
QY 53 SerlyspProValArgleuAspAsnIleleu-----GlnIleArgHispThr 67  
Db 30313 AACATGTTGTCAGTTTCAGAGGCTCTCATCTCAGCCATCAACGCCCAAGCATCATCATC 30254

QY 68 ProHistrhAsnGlyLeuSerAspIleTyrleuLeuAsnAspProHIsGluAlaPheAla 87  
Db 30253 CCCAATCAATCTGGA-----TATTATCCGATTTACAAACAGGTGCCAATGCGTTGCT 30203  
QY 88 AlaTrgAlaAlaIleuIleGluSerAlaGlnHIsSerleuAspLeuGlnTyrIleTirp 107  
Db 30202 GCACGAGTACGCTGACTGATGAGTGCAACCAACCAATCATATTCATATCATATTGG 30143  
QY 108 ArgAsnAspIleSerGlyArgleuLeuPheAsnleuValTyrleuAlaAlaIleArgGly 127  
Db 30142 CATAAACGATGAGCTGCGTCACTGATGCTTAAACCGCTGGGAACGCGTCACTGCTG 30083  
QY 128 ValArgValArgleuLeuAspAsp--AsnAsnThrArgGlyLeuAspAspLeu 146  
Db 30082 GTGATTGTCATGCTGCTGATGATGATTTATAGCTACCTGACCTGATCATCTCTTA 30023  
QY 147 LeuAlaLeuAspSerHisProAsnIleGluValArgleuPheAsnProPheValLeuArg 166  
Db 30022 CTACGATCTCAAGACATAAATAATGCTGATGATTAATCAATCCATGCTTATCTG 29963  
QY 167 LysTrpArgAlaLeuGlyTyrleuThrAspPheProArgleuAsnArgMetHisAsn 186  
Db 29962 GCGTCCGCTAGCTCAACTATATGTTACATCCATACCGACCATCGCGCTATCATAT 29903  
QY 187 LysSerPheThrAlaAspAsnArgAlaThrIleleuGlyArgAsnIleGlyAspGlu 206  
Db 29902 AAAAGCATGACATTTGATTAATAAATCATAGTGTAGTGTGCTATATCGTAAAGAA 29843  
QY 207 TyrPheLysValGlyLysPThrValPheAlaAspLeuAspIleAlaThrGlySer 226  
Db 29842 TATTTAATATGACACCTGTTAATCATTTGCTATTTGGATGTCATGTTAGTGGTCAT 29783  
QY 227 ValValGlyLysValSerHisAspPheAspArgTyrTrpAlaSerHIsSerAlaHisAsn 246  
Db 29782 GTCGTGCAAAATTTACCAAAAGCTTGAGATTATGGGCTTACCCCTATCCTTTCAT 29723  
QY 247 AlaThrArgIleIleArgSerGlyAsnIleGly-----LysGlyLeuGlnAlaLeuGly 264  
Db 29722 ATTGAAACACTTGTCAAAACATGACACCAAGATGATCATGCTAGTCATCAADCTGTGTA 29663  
QY 265 TyrAsnAsp-----GlnThrSerArgHIsAlaLeuAlaArgTyr 277  
Db 29662 TTTTATGACTTGAAAAAGTAAAGACAGCTCACTGATGCTGACCTTACGCACTTAT 29603  
QY 278 ArgGluThrValGlnGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTyr 297  
Db 29602 CGCCAGCATGCAAAATATCAACATGCGCAAGATTACTCGCCAGCAAGTACCATTT 29543  
QY 298 GlnSerValGlnThrArgleuIleSerAspAspProAlaLysGlyLeuAspArgAspArg 317  
Db 29542 TTTTGACAGAGATTGATTGATGACAGATTAATTCAAATTAACGGTCACTCAGAT 29483  
QY 318 ArgLysProProIleAlaGlyArgleuGlnAspAlaLeuLysGlnProGlyLysSerVal 337  
Db 29482 CCTAATGATTTTGGTTTCACACTTACAAACACATTTGGTCAACCAAGAAAAATC 29423  
QY 338 TyrleuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeu 357  
Db 29422 AGCATCATCTCTCTTATTTATTTGTTCTACCAAAAGAGTGTGCAACACTGATCACCCTT 29363  
QY 358 ValGlnAspGlyIleAspValThrValleuThrAsnSerleuGlnAlaThrAspValAla 377  
Db 29362 GCTAAATGGTGTGAGTGCATATTTTGACCAATTCATTTGATGCCAGGATGGTGT 29303  
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProleuLeuLysAlaGlyIleLysLeu 397  
Db 29302 ATTTGACTGACAGTTTACGCACTGCGGAAGCAACTACTGCGGAGCGGTGCATCTT 29243  
QY 398 TyrIleuGlnProAsn--HisAlaValProAlaThrIleLysAspLys-----GlyLeu 414  
Db 29242 TTTGAGATTAAATCATAGTGTGATGATCCAGATTAACGAAATCGTTTGTGCGTACA 29183  
QY 415 ThrGlySerSerValThrSerleuHisAlaLysThrPheIleValAspGlyLysArgIle 434



```

Db 634 GTACCCAGCAGCGTGGGGGACAGCTTGCACAGTGAACACAGCCCGCCAGCCGCGG 693
Oy 247 AlArghIleIleArghSerGIYAsnIIleGIYsGIleuGlnAlaLeuGIYTrAsn 266
Db 694 ATCACCCGATTCCTCGCATGACCGGATCCAGCAACGACTGCGGCTACGCCAGCCG 753
Oy 267 AspGIuThrSerArghHsAlaLeuLeuArghTrArghIleuValGlnIleuSerProleu 286
Db 754 CTGGAAGTGTG-----CTGGCCAAAGCCAGG-----ACCCAGCCGCAAGGCCCTG 798
Oy 287 TyrGIuIleGIleuIleu-----GIYArghIleAsp-----TyrGIuSer----- 299
Db 799 TACAGCCGCTGTGATGCGCTATCACTCGACGCGCTGTGATGTGTGGCGCAAGCAGCTG 858
Oy 300 -----ValGIuThrArghIleuSerAspAspProAlaIleuGIYLeuAspArgh 317
Db 859 ATCTGGGCTGCAGCAGAGCCCTATGGATGCGCCAGCAAGGTGTGTGGCAAGACGAA 918
Oy 318 ArghIleProPro-----IleAlaGIYArghIleuGlnAspAlaLeuGIYsGIuProGIuIle 335
Db 919 CCGGACCCGCAATTACTGTGAGCAGCACTGCTCGGACCTCGGCTAACGTCACCGC 978
Oy 336 SerValTyrIleuValSerProTyrPheValProThrIleuSerGIYTrAspAlaLeu 355
Db 979 GAGTGGCTGCTGGCATGCGCTACTTTGTACCGGCGGCAAGCGGCTGCTTACCTGACC 1038
Oy 356 LysLeuValGlnAspGIYIleAspValIleValLeuThrAsnSerIleuGlnAlaThrAsp 375
Db 1039 GGCGCGCCGCGAGCTGCGTTCGTGAGCTGTGACCAACTGCTGAGCAAGCCAGCAGC 1098
Oy 376 ValAlaAlaValHisSerGIYTrValIleYrArghIleuProleuGIYsAlaGIYIle 395
Db 1099 GTGCGGCGAGCTGTGAGGCTATGACCCCTATGCTGTGCTTGTGACGAGCGGCTG 1158
Oy 396 LysLeuTyrGIleuGlnIleuProAsnHsAlaValProAlaThrIleuAspGIYLeuThr 415
Db 1159 CAGCTTATACAGACTTCCGCGCCAGCCAGCCAGCCAGCGGCGGCGCACTGATTCGCT 1218
Oy 416 GYSerSerValIleThrSerLeuHsAlaIleuThrPheIleValAspGIYsArghIlePhe 435
Db 1219 GCGAGCTGCGTGTGAGCTGTGACCAAGCGATGCTGTGCGACCGCGCAAGACCTTC 1278
Oy 436 IleGIuSerPheAsnLeuAspProArghSerAlaIleuGlnIleuGlnIleuVal 455
Db 1279 ATGCGCTGTCAACTGTGACCGCGCTCAGTATTGTGGAACCGAGGTGGGCTACTG 1338
Oy 456 IleGIuSerProIleuIleuGlnIleuGlnIleuGlnIleuAsp-----ThrSer 473
Db 1339 GTGGACAGCCGCGAGCTGTGGCGAA---TACACCGCGCAACTGCGCCAGCAGGATAGCG 1395
Oy 474 ProGIuTyrAlaIleuArghValIleuLeuAspArghHsAsnArghIleuGlnIleuPhe 491
Db 1396 CCGTGGCTGTGATGAGTGAACCTG---ATAGGTAACAAGCTGTGTGGCGCAGCAGC 1452
Oy 492 AspProAlaThrArghIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 511
Db 1453 GACAAATGCCAAGCAGCGCTGCTGAGCTGTGAGCCG---GGCGGAGATCTGGCGAGCTTC 1509
Oy 512 AlaAlaIleuIleuSerIleuLeuProIleuIleuSerIleuLeu 525
Db 1510 AATGCCGTGATCAGCAGGCGGTGTGCTGTGAGAAATGTGA 1551
RESULT 8
AAV24861
ID AAV24861 standard; DNA; 1509 BP.
AC AAV24861;
XX
XX
XX 24-JUN-1998 (first entry)
DE H. pylori secreted protein ORF 02ae11612_33203250_c1_51.
XX

```

```

KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 1..1509
FT /tag= a
PN MO9737044-AL.
PD 09-OCT-1997.
PE 27-MAR-1997; 97MO-US05223.
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.
PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
PA (ASTR ) ASTRA AB.
XX
XX Alm RA, Smith D;
PI
XX
XX WPI; 1997-503122/46.
DR P-PSDB; AAW55452.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
PS Claims 5,6,37; Page 290; 1145pp; English.
XX
XX This sequence encodes a H. pylori secreted protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds. The
XX useful as potential H. pylori life cycle activators or inhibitors. The
XX DNA and probes derived from it may be used for the identification of
XX H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
XX acid sequences complementary to the DNA act as antisense sequences and
XX can be used to prevent the translation of H. pylori mRNA. Antibodies
XX against the protein can be used in immunoassays to evaluate the abundance
XX and distribution of H. pylori-specific antigens. The genomic sequence of
XX H. pylori (ATCC 55679) was determined from overlapping contigs generated
XX by mechanically shearing the bacterial DNA. The sequences were analysed
XX for ORF of at least 180 nucleotides, and the predicted coding regions
XX defined by computer evaluation. To identify likely H. pylori antigens for
XX vaccine development, the amino acid sequences predicted from various ORF
XX were analysed for significant homology to other known or exported
XX membrane proteins. Having identified and determined the sequences of
XX interest, particular regions can be isolated from H. pylori by PCR
XX amplification for recombinant polypeptide production, e.g. in E. coli
XX hosts.
SQ Sequence 1509 BP; 471 A; 302 C; 277 G; 459 T; 0 other;
Alignment Scores:
Pred. No.: 5,1e-55 Length: 1509
Score: 632.00 Matches: 163
Percent Similarity: 51.52% Conservative: 92
Best Local Similarity: 32.93% Mismatches: 184
Query Match: 23.30% *Indels: 56
DB: 18 Gaps: 17
US-10-066-551-4 (1-525) x AAV24861 (1-1509)
Oy 64 IleArghIleIleArghIleuSerAspIleuIleuIleuIleuIleuIleuIleu 77
Db 88 ATCTCTATGATACCCCTTACACTACCACTTGTGATGCTTAATAATTAAAGAA 147
Oy 78 -----LeuLeuAsnAspProHsIleuGlnAlaIleuArgh 89

```



CC to the DNA act as antisense sequences and can be used to prevent the  
 CC translation of H. pylori mRNA. Antibodies against the protein can be  
 CC used in immunoassays to evaluate the abundance and distribution of  
 CC H. pylori-specific antigens. The genomic sequence of H. pylori  
 CC (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.

XX Sequence 1509 BP; 471 A; 302 C; 276 G; 460 T; 0 other;

# Alignment Scores:

Pred. No.:	51e-55	Length:	1509
Score:	632.00	Matches:	163
Percent Similarity:	51.52%	Conservative:	92
Best Local Similarity:	32.93%	Mismatches:	184
Query Match:	23.30%	Indels:	56
		Gaps:	17

US-10-066-551-4 (1-525) x AAHV24959 (1-1509)

QY 64 ILeaRgHISrHrProHISrHrAsnGlyLeuSerIleTyR----- 77  
 DB 88 ATCTCTATGACCCCTTACACTACACCATGTGGAGCTGTATGCTAAATTTAAAGAA 147  
 QY 78 -----LeuLsnaSPrroHISrHrAsnGlyLeuSerIleTyR 89  
 DB 148 AACCTTAACATAGTGGCCCATCTTTTAAAGACGCTTGTGATGCTTGTGCACGAA 207  
 QY 90 AAlaLeuIleGlySerAlaGluHisSerLeuAspLeuGlnTyRTrIleTrrpArgsn 109  
 DB 208 GTGGGTCTATTAGATGAGCAAAAGCAATGTGCATGCAAACTTATATCTATAAAG 267  
 QY 110 AspIleSerGlyLeuLeuPheAsnLeuValTyRLeuAlaIleGluArgGlyValArg 129  
 DB 268 GACCTTCTCTCAAGTATGCTTAAGACACTTTTAAAGCCGCAATGTGGGTAAAA 327  
 QY 130 ValAlaLeuLeuLeuAspAsnAsnThrArgGlyLeuAsp-----AspLeu 145  
 DB 328 GTGGCATTCTTTAGATGACAAAC-----GGGTAGATGATGATTTTCAGACATC 378  
 QY 146 LeuLeuAlaLeuAspSerHisProAsnIleGlyValArgLeuPheAsnProPheValLeu 165  
 DB 379 ATGCTT---TTAAATTTCCACAAAACATGTAAGTAAGAAATCTTTAACCCCTACTATATC 435  
 QY 166 ArgLysTrpArgAlaLeuGlyTyR-----LeuThrAspPheProArgLeuAsnArg 182  
 DB 436 CGCAAT---AAAGGCTTGGCTTATTTGAAATGCTTGGCGGATTAAGACGATTAAGAAA 492  
 QY 183 ArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaTrIleLeuGlyValArgsn 202  
 DB 493 CGCATGCAACAAAGCTTTTATGCTGATATTTGCTGTCATATTAATGAGGGGGGCAAC 552  
 QY 203 IleGlyAspGlyTrpPheLysValGlyGluAspThrValPheAlaAspLeuAspIleLeu 222  
 DB 553 ATTGCGGACATATTATTGATTAAGCATTTTACACGAAATTTTATGATTTAGACCTTGG 612  
 QY 223 AlaThrGlySerValIleGlyValSerHisAspPheAspArgTyRTrpAlaSerHis 242  
 DB 613 TTTTGGGGGGTGGTTCACAAAGGCCAAAGACTTTGAAATTAATTTGGCGATTCAC 672  
 QY 243 SerLHisAsnAlaThrArgIleLeuArgSer-----Gly 254  
 DB 673 CGCTATCTCCTGTTTCA---TTATTAAGAACCATTAAGACTCAAAACACGTTTAA 729  
 QY 255 AsnIleGlyLysGlyLeuGlnAlaLeuGlyTyRAsnAspGlyTrpSerArgHisAlaLeu 274

DB 730 GAATGCTGTAAGCTTCATGAAATAATCCCTATACAGCGCTGAAGACGCAATTTGAA 789  
 QY 275 LeuArgTyRArgGluThrValGluInSerProLeuTyRgln---LysIleGlnThrGly 293  
 DB 790 AAAAAGTCATGATTTTATAGAACGCTTCCAAAATATCCAAATACCATTTATATATGG 849  
 QY 294 ArgIleAspTrpGlnSerValIleThrArgLeuIleSerAspProAlaLysGlyLeu 313  
 DB 850 -----AATGCCATTTTCTACAGCGATTTGCCGCCAAA---ATY 885  
 QY 314 AspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnPro 333  
 DB 886 GACACGCCCTTGATTCG---CCCATCAAAATCCGCTTTGAAAGACCTTAAGAACCT 942  
 QY 334 GluLysSerValTyRLeuValSerProTyRPhenValProHlySerGlyThrAspAla 353  
 DB 943 AAGACTCCGTTTTATGCGTTCAATGATTTTATCCAGGTAAGATAATGAATC 1002  
 QY 354 LeuAlaLysLeuValGlnAspGlyLysAspValThrValLeuThrAsnSerLeuGlnAla 373  
 DB 1003 TTTAAATAATCAAAATTTCTAAGGGGATGAAATGGAATATTTTACCAATTTCCCTTCATCT 1062  
 QY 374 ThrAspValAlaIleValHisSerGlyTyRValLysTyRArgLysProLeuLysAla 393  
 DB 1063 ACGACGCTATCGTGTGTATGCGGCGTGGCAAGGTATCCCATTAATTTATGCGCAATC 1122  
 QY 394 GlyIleLysLeuTyRglnLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly 413  
 DB 1123 GCGCGGATGCTTATGAAATACGAATTTTTCACACCCGCAATTAAGAGCGCG--- 1179  
 QY 414 LeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArg 433  
 DB 1180 -----TTTACACCAACATCTTACACGCGCAAGACAAATGTTTGAACGATGCCCTA 1233  
 QY 434 IlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMetGly 453  
 DB 1234 ACGCTCTAAGGAGTTTAAATATGATGCGGCGTCTGATATTAACACGTAAGAGTGG 1293  
 QY 454 ValValIleGlySerProLysIleAlaGluInMetGluArgThrLeuAlaAspThrSer 473  
 DB 1294 GCTTGTGTTGCAACCCCTTTTGGCCAAAGGGGCTTGTGTGCTTAAAGATCAATGCC 1353  
 QY 474 ProGluTyRAlaTyRArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspPro 493  
 DB 1354 ---CAACATCATGCGATTTAGTCTTGTATGCGCAT---AGAGTGATTTGGGAA----- 1401  
 QY 494 AlaThrArgLysThrTyR-----ProAsnGlnProGluAlaLysLeuTrpLys 509  
 DB 1402 GCTACAGAAAGAGCATCTTAATCCATGAATAAAACCTGCTGATACCTCTTCTTTTGG 1461  
 QY 510 ArgIleAlaLysIleLeuSerLeuLeuProIleGlnSerLeu 524  
 DB 1462 CGTTTGATTAAGATGCTTAAGTCTCTCTGTAAGAGAACTT 1506  
 RESULT 10  
 ID AAX14097 standard; DNA; 1630 BP.  
 AC AAX14097;  
 DT 31-MAR-1999 (first entry)  
 DE H. pylori GHPD 1375 gene.  
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KW peptic ulcer disease; ss.  
 OS Helicobacter pylori.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 294..1580  
 FT /\*tag= a



PN M09843478-A1.  
 XX 08-OCT-1998.  
 PD  
 XX 01-APR-1998; 98WO-US06371.  
 XX 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
 DR WPI: 1998-542293/46.  
 DR P-PSDB: AAW98378.  
 XX  
 PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases  
 PS Claim 1; Page 581-583; 2054pp: English.  
 XX  
 CC This sequence represents a polynucleotide of the invention. It was  
 CC isolated from Helicobacter pylori and encodes a H.pylori GHP protein.  
 CC The polypeptides can be used for preventing or treating Helicobacter  
 CC infections, and gastrointestinal diseases associated with these  
 CC infections, including acute, chronic, and atrophic gastritis, and peptic  
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used  
 CC for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.  
 XX  
 SO Sequence 1630 BP; 493 A; 310 C; 318 G; 509 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 1,47e-54 Length: 1630  
 Score: 628.00 Matches: 163  
 Percent Similarity: 48.85% Conservative: 93  
 Best Local Similarity: 31.11% Mismatches: 170  
 Query Match: 23.15% Indels: 98  
 Gaps: 19  
 US-10-066-551-4 (1-525) x AAX14097 (1-1630)  
 Oy 52 ThrSerLysProValArgLeuAspAsnIleuGlnIleArgHisThrProHisThrAsn 71  
 Db 147 TCAGCTCTCCATCTCTTATGAT-----CCCTACACTACC 182  
 Oy 72 GlyLeuSerAspIleTyr----- 77  
 Db 183 CCCATTGGAGCTGTATGCTGAAATAATTAAAGAAACCTAACCATAGCGCGCATAT 242  
 Oy 78 LeuLeuAsnAspProHisGlnAlaIleAlaIleAlaIleAlaIleIleGlnSerIleGln 97  
 Db 243 CTTTAGAAGATGGCTTGGCTGTGTCATAGAGTGCGCATTTATTAAGTAAGCGCA 302  
 Oy 98 HisSerLeuAspLeuGlnIleTyrIleThrParGAsnAspIleSerGlyArgLeuLeuPhe 117  
 Db 303 AAAACATTCATGACCTAATATATATATAAAACGACCTTCTTCACAGTGATGCT 362  
 Oy 118 AsnLeuValTyrLeuAlaIleGlnAlaIleArgValArgLeuLeuLeuAspAsn 137  
 Db 363 AAAGAACTTTAAATGCGCCAACTGCGGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 422  
 Oy 138 AsnThrArgGlyLeuAsp-----AspLeuLeuLeuAlaLeuAspSerHisPro 153  
 Db 423 -----GGATTGCTATTCGATTTCGATTTCGATTTCGATTTCGATTTCGATTTC 470  
 Oy 154 AsnIleGluValArgLeuPheAsnProPheValLeuArgGlySerThrParGAlaLeuGln 173  
 Db 471 AACATTGAGTGCAAAATTTTAAACCCCTACTATATCCGCAAT---AAAGCTTGGCTTAT 527

Oy 174 -----LeuThrAspPheProArgLeuAsnArgMetHisAsnLysSerPheThr 190  
 Db 528 TTTGAATGCTTGGCGGATTAAGCGCATTAATAAAACCATGCACAAACGCTTTTCATC 587  
 Oy 191 AlaAspAsnArgAlaThrIleLeuGlnIleArgAsnIleGlyAspGlnTyrPheLysVal 210  
 Db 588 GTGATTAATTTCCGCTGCTATAGGGGCGCAATTAATTTGGGCAATTAATTTGATTAAC 647  
 Oy 211 GlyLysAspThrValPheAlaAspLeuAspIleLeuAlaIleThrGlySerValValGly 230  
 Db 648 GATTTAGACAGCAATTTTTCATTTAGACGCTTTGTTGGGGGCTTCTCAAA 707  
 Oy 231 ValSerHisAspPheAspArgTyrThrParLaseHisSerAlaHisAsnAlaThrArgIle 250  
 Db 708 GCCAAGAAAGCTTTGAACGCTATTGAGATTCACACGCTCATGCCGTTTCA---TTA 764  
 Oy 251 IleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGlnThrSer 270  
 Db 765 CTAAAGAAC----- 773  
 Oy 271 ArgHisAlaIleLeuArgTyrArgIleThrValGlnGlnSerProLeuTyrGlnLysIle 290  
 Db 774 ---CATAAAGACTCAAA---AACAGCTAAAGAAATCGTAACCTCATTAATAAATC 827  
 Oy 291 GlnThrGlyArgIleAspThrPheIleSerValGlnThrArgLeu----- 304  
 Db 828 CCTATCAGCGCTGAGACAAACAAACAGTTTGAAACAAAGCAATGATTTTATAGTGT 887  
 Oy 305 -----IleSerAspAspPro 309  
 Db 888 TTCCAAATAATCCAAATACCCCATTTTATGGAATGCCATTTTGGCCATTCACCC 947  
 Oy 310 AlaLysGlyLeuAspArgAspArgArgLysProPheIleAlaGlyArgLeuGlnAspAla 329  
 Db 948 AAAAAA---ATTGACACGCCCTTGATTCG---CCTATCAAAATGCGCTTTGAGAAAGCC 1001  
 Oy 330 LeuLysGlnProGlnLysSerValTyrLeuValSerProTyrPheValProThrLysSer 349  
 Db 1002 CTTAAACAGCTAAGACGCTGCTTTTATGCGCTGATGATTTATTCACAGCAAAAG 1061  
 Oy 350 GlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsn 369  
 Db 1062 ATGATGAATAATCTTTAAATCAATTTCTAAGGGGATTAATGATCACTTACCAAT 1121  
 Oy 370 SerLeuGlnAlaThrAspValAlaValHisSerGlyTyrValLysTyrArgLysPro 389  
 Db 1122 TCCCTTTGATCTAGTGAATGCGATGCTATGAGGCGATGGAAGATGCAACCA 1181  
 Oy 390 LeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaValProAlaThr 409  
 Db 1182 TTAGTGCAGATGCGCGCATGTCTATGAAATACGAAACGATTTTTCACCGCAGAT 1241  
 Oy 410 LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleVal 429  
 Db 1242 AAAGGCGC-----TTTAGCACCAAACTCTCTGCAAGGCAAGCGATTTGTTT 1292  
 Oy 430 AspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsn 449  
 Db 1293 GATGACATTTTAACGCTCTAGAGATTTCAATATGATCCGCGCTGCAATACATCAAC 1352  
 Oy 450 ThrGlnMetGlyValValIleGlnSerProLysIleAlaGlnIleMetGlnArgThrLeu 469  
 Db 1353 ACTGAAGAGCGGCTTTGTTGACAAACCGCTTTTGAAGAGCGTGGCTTGGCTT 1412  
 Oy 470 AlaAspThrSerProGlnTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGln 489  
 Db 1413 AAAGATCATGCC---CAACATCATGCAATTGGTGCGGTATACGGCAT---AGAGTGAT 1466  
 Oy 490 Trp-----HisAspProAlaThrArgLysThrTyrPro 500  
 Db 1467 TCGAAGCGGTGGAAGAGCATTTTAATCATGAA-----AAAGCT----- 1508  
 Oy 501 AsnGluProGlnAlaLysLeuThrLysArgIleAlaIleLysIleLeuSerLeuLeuPro 520

```

Db 1509 ---TCGGCTGACACTTCCTCTTTTGGCGTTGATTAAAGATGCTTAAGCTCTTCT 1565
Qy 521 IleGlusertleu 524
Db 1566 GAAAGAGAGCTT 1577

RESULT 11
AB069041
ID AB069041 standard; DNA; 2319 BP.
XX
XX AB069041:
AC
XX 29-AUG-2002 (first entry)
DE
XX Listeria monocytogenes 4b contig DNA sequence #1807.
DE
XX
XX Antibacterial; Listeria; food contamination; mutational analysis;
KM infection; ds.
XX
XX OS Listeria monocytogenes 4b.
XX
XX PN MO200228891-A2.
XX
XX PD 11-Apr-2002.
XX
XX PF 04-OCT-2001; 2001WO-FR03061.
XX
XX PR 04-OCT-2000; 2000FR-0012697.
XX
XX PA (INSP ) INST PASTEUR.
XX
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX PI Kunst F, Glaser P;
XX
XX MP: 2002-332479/37.
XX
XX PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
XX PS Claim 14; SEQ ID 1854; 180bp; French.
XX
XX CC The present invention relates to nucleic acid sequences
XX CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
XX CC and primers for identification and/or detection of Listeria (e.g. as
XX CC contaminants in foods, or mutational analysis) and for analysis of
XX CC gene expression. Proteins encoded by the nucleic acid sequences can be
XX CC used to screen for compounds that modulate gene expression, replication
XX CC and pathogenicity of Listeria (potential therapeutic agents), also for
XX CC treating infections by Listeria, and are useful as immunogens in
XX CC anti-Listeria vaccines.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 2319 BP; 759 A; 368 C; 489 G; 702 T; 1 other:

Alignment Scores:
Pred. No.: 6,59e-23 Length: 2319
Score: 323.00 Matches: 124
Percent Similarity: 37.93% Conservative: 74
Best Local Similarity: 23.75% Mismatches: 214
Query Match: 11.91% Indels: 110
DB: 24 Gaps: 17

US-10-066-551-4 (1-525) x AB069041 (1-2319)
Qy 4 AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23
Db 555 AATCCACGACCTAGAAAATTC-----AGCACAGCTCAAGTAATGAAGAAAAGCTAAATTA 608
Qy 24 IleSerLeuLeuGlyLeuLeuLeuGlySerSerSerIleProLeuProLeuGluGlu 43

```

```

Db 609 ATTATGGGATTTCACGCAATTCACAAACACACTATGAAAAATTAACACGACTATCTAAA 668
Qy 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
Db 669 AGAATT-----GGCATTTTACGCTCCATAGAGCCAAATTAG---GGCAATTAATGAA 719
Qy 64 IleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnProHis 83
Db 720 ATATTA-----ACAAATGGC-----GAA 737
Qy 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGln 103
Db 738 GAAACTTTTCCAGTCTCTTAGACGCGCTTAGAAAGCTGGAACACACATTCATATTCAA 797
Qy 104 TyrTyrIleTrpArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAla 123
Db 798 TACTATATTTTCAAAACGAGATGAAATTTCTCAAAAGATTCGTATTTTGGTAGAAAA 857
Qy 124 AlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArgGlyLeuAsp 143
Db 858 GCAAAACGACGCGTTGAAGTAGATTATTTATGATGGCTTGGATCAGCAAACTGGGT 917
Qy 144 AspleuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 918 AAAGCTTTTATGCTCTTGAAGAAAGCTGAGCTGATGATTCACGATTTGACCA--- 974
Qy 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183
Db 975 ATTACCTCTCCCTGGATTGTA-----AGAACGCAATTTAGA 1013
Qy 184 MethiAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIle 203
Db 1014 AATCACCGTAATAATGTCGTGATGATGATGATGATGATGATGATGATGATGATGAT 1073
Qy 204 GlyAspGluTyr-----PheLysValGlyLysAspThrValPheAla 217
Db 1074 GGACAGAGATATTCCTTATATACACAGATTTCCGTGTTGGCGGATACA----- 1124
Qy 218 AspleuAspIleLeuAlaThrGlySerValValGlyLysValSerHisAspPheAspArg 237
Db 1125 -----CATATTAATACACAGCGCAAGCTGTATACAGCTCAAGATCTTTCTAAT 1178
Qy 238 TyrTrpAla-----SerHisSerAlaHisAsnAlaThrArgIleLeuArgSerGlyAsn 255
Db 1179 GATTGGGCTATATGAGAAACCAAGCGTGCCTCAATATTAATTAATGAGAGCTGA 1238
Qy 256 IleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeu 275
Db 1239 TTACAAACAATATTTTTCACACGTTGATGTGGGTACGAAATGGCG----- 1283
Qy 276 ArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIle 295
Db 1284 -----CAAGTATTTACGGCGGACCATAT 1307
Qy 296 AspTrpGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315
Db 1308 GAT----- 1310
Qy 316 AspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLys 335
Db 1311 -----AAAGAAAATAGGTTCCGCGATTCGATTCGATTAATTTACCGCTAAAGAA 1364
Qy 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355
Db 1365 TCTGTTGGATTTGTGCGCTTACTTGTTCACATAGAGAGTGGCTGCAGTATATTCGC 1424
Qy 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375
Db 1425 CCGGTTCGATGACGCGTGTGACGTCGAGTTATTTTCACAGC-----AAAGTGAT 1478
Qy 376 ValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIle 395

```

Db 1479 CGGGGATTCATTCACGGAAGTAATCGCATGTGAAACAGATTGAACGAGCGCG 1538  
 Qy 396 LysLeuTYRGIuLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThr 415  
 Db 1539 AAAATGTATGCTTATGCGGATGATCTTTGTT----- 1571  
 Qy 416 GlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePhe 435  
 Db 1572 -----CATGCGAAGCAATGATGAGCGAGCGGAGCGGTCAGCT 1610  
 Qy 436 IleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValVal 455  
 Db 1611 ATAGGTACTGCCAATTTGACGTGCGTAGTTTATGATTAATCATCAATTAATGATATTC 1670  
 Qy 456 IleGlySerProLysIleAlaGluGlnMetGluArgThrIleuAlaAspThrSerProGlu 475  
 Db 1671 TTATATGATGAA-----AGCGAGGCTATGCATCTTTAAACGCGATTTTGAGAAAGAT 1724  
 Qy 476 TyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThr 495  
 Db 1725 TTT-----GAAGATAGCCGACTT-----TTTACG 1748  
 Qy 496 ArgLysThrTyrProAsnGlnProGluAlaLysLeuThrPheLysArgIleAlaLysIle 515  
 Db 1749 ATGAAAGATGAGAAACAAACATTATGACTCGTATAAAGAGTTCTATCTAGTTTA 1808  
 Qy 516 LeuSer 517  
 Db 1809 CTATCA 1814  
 RESULT 12  
 ID ABQ71029 standard; DNA; 6273 BP.  
 AC ABQ71029;  
 DT 29-AUG-2002 (first entry)  
 XX Listeria monocytogenes 4b contig DNA sequence #971.  
 DE Listeria monocytogenes 4b contig DNA sequence #971.  
 XX Antibacterial; Listeria; food contamination; mutational analysis;  
 KW Infection; ds.  
 XX Listeria monocytogenes 4b.  
 OS WO200228891-A2.  
 PN 11-APR-2002.  
 PD 04-OCT-2001; 2001WO-FR03061.  
 PF 04-OCT-2001; 2000FR-0012697.  
 PR 04-OCT-2000; 2000FR-0012697.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PI Kunst F, Glaser P;  
 XX WPL; 2002-332479/37.  
 DR New genomic sequences from Listeria species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PS antibodies and modulators  
 XX Claim 14; SEQ ID 3842; 180pp; French.  
 CC The present invention relates to nucleic acid sequences  
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of Listeria (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of Listeria (potential therapeutic agents), also for

CC treating infections by Listeria, and are useful as immunogens in  
 CC anti-Listeria vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from Wipo at [http://wipo.int/publ/published\\_pct\\_sequences](http://wipo.int/publ/published_pct_sequences).  
 XX  
 SQ Sequence 6273 BP; 2064 A; 1047 C; 1296 G; 1855 T; 11 other;  
 Alignment Scores:  
 Pred. No.: 2,85e-22 Length: 6273  
 Score: 323.00 Matches: 124  
 Percent Similarity: 37.93% Conservatave: 74  
 Best Local Similarity: 23.75% Mismatches: 214  
 Query Match: 11.91% Indels: 110  
 DB: Gaps: 17  
 US-10-066-551-4 (1-525) x ABQ71029 (1-6273)  
 Qy 4 AsnProLysThrGlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeu 23  
 Db 555 ATCCAGACAGTAAAGAAATTC-----AGCACAGCTCAAGTAAATGAAACCTAAATTA 608  
 Qy 24 IleSerLeuLeuCysLeuLeuLeuLeuCysSerCysSerSerTyrLeuProProLeuGluGlu 43  
 Db 609 ATTAATGGGATTCACGCAATTCACAAACAACTTAATGAAATTAACACGACTATCTAA 668  
 Qy 44 ArgThrGlySerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63  
 Db 669 AGAATT-----GGCCATTTAAAGCTCATAGAGCCCAATTAAG--GGGATTAATAATCGAA 719  
 Qy 64 IleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHis 83  
 Db 720 ATATTA-----ACAAATGGC-----GAA 737  
 Qy 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGln 103  
 Db 738 GAACCTTTTCAGTCGCTCTTACGCGCGCTTAGAAGGCGTAAAGGCAACACATTCATATCA 797  
 Qy 104 TyrTyrIleTyrArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAla 123  
 Db 798 TACTATATTTTCMAACGCGATGAATTTCTACAAAGATTGCTGATATTTTGGTAGAAAAA 857  
 Qy 124 AlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThrArgGlyLeuAsp 143  
 Db 858 GCAAAAGCAGCGGCTGAGATTGATTGTTGAGTGGGCTTGATCAAGCAAACTGGGT 917  
 Qy 144 AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163  
 Db 918 AAAGCCTTTTAGCTCTTGAAGAGCGTGAAGTATTCACGCAATTTGACCA--- 974  
 Qy 164 ValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183  
 Db 975 ATTACCTCTCCCTGGATTTGA-----AGAACAGCCCAATTTAAGA 1013  
 Qy 184 MetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIle 203  
 Db 1014 AATCACCGTAAATATGCTGATGATGATGCTCAATTTGGCTTAGACGGCGGCTTAATAT 1073  
 Qy 204 GlyAspGluTyr-----PheLysValGlyGluAspThrValAla 217  
 Db 1074 GGAGAGAAATATGCTTCAATACACAGATTTCCCTGTTGGCGGATACA----- 1124  
 Qy 218 AspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAspPheAspArg 237  
 Db 1125 -----CATTTTAAATATACAGCCAGCTGTATTCGAATCTCAAGATCTTTTSTAAT 1178  
 Qy 238 TyrTrpAla-----SerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsn 255  
 Db 1179 GATTGGCTATATGAGAAACCAACGACGATGCTGACAGATTAATTAATGACAACTGGA 1238  
 Qy 256 IleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeu 275  
 Db 1239 TTACACAAATATTTTTCACGAGTGTGATGTGGTGCAGAAATGGCG----- 1283

```

QY 276 ArgTyrArgGluThrValGluGlnSerProLeuTyrGlnIleGlnThrGlyArgIle 295
DB 1284 -----CAAGTATTTACGGCGACCATAT 1307
QY 296 AspTrpGlnSerValGlnThrArgLeuIleSerAspSerProAlaLysGlyLeuAspArg 315
DB 1308 GAT----- 1310
QY 316 AspArgArgLysProPheIleAlaGlyArgGlnAspAlaLeuLysGlnProGluLys 335
DB 1311 -----AAGCAAAATGGGTTCCGGATTCGATCGCTGATTAAATTGATCCGGCTAAAGAA 1364
QY 336 SerValIleLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355
DB 1365 TCTGTGGATTTGTCGCCCTTACTTGTCCAGATGAGAGTGGCTGACAGTATTCCG 1424
QY 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375
DB 1425 CGGGTTCGATGACCGGTGTGACGTGCGAGTATTTATTCAGGC-----AAGGTGAT 1478
QY 376 ValAlaAlaValHisSerGlyTyrValIleTyrArgLysProLeuLeuLysAlaGlyIle 395
DB 1479 CGCGGATTCATTCACGGAAGTAAGCGATATGTGAAACAAATGATTTGAGACGCGCG 1538
QY 396 LysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThr 415
DB 1539 AAATGTATGCTTATGCCGATGATCTTTGTT----- 1571
QY 416 GlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePhe 435
DB 1572 -----CATGCAAGGCAATGCTAGTGACGCGACCGCGACGCT 1610
QY 436 IleGlySerPheAsnLeuAspProArgSerValArgLeuAsnThrGluMetGlyValVal 455
DB 1611 ATAGTACTGGCATTTTGCAGCTGCTAGTTTATGATTGATCATGAAATATATGATATTC 1670
QY 456 IleGluSerProLysIleAlaGluGlnMetGluArgThrLeuAlaAspThrSerProGlu 475
DB 1671 TTATATGATGAA-----AGCGAGGCTATGATCATTTAAACGATTTTGAGAAAGAT 1724
QY 476 TyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrAspProAlaThr 495
DB 1725 TTT-----GAAGATAGCGGACTT-----TTTACG 1748
QY 496 ArgLysThrTyrProAsnGlnProGluAlaLysLeuTyrLysArgIleAlaLysIle 515
DB 1749 ATGAAGAATATGAAACAAACCATTTATGACTGCTATTAAGAAAGTTCTATCTAGTTTA 1808
QY 516 LeuSer 517
DB 1809 CTATCA 1814

RESULT 13
AB067196/c
ID AB067196 standard; DNA; 684707 BP.
XX
AC AB067196;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #9.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001MO-FR03061.

```

```

XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
PI Kunst F, Glaser P;
PI
DR WPI: 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
PS
PS Claim 5; SEQ ID 9; 180pp; French.
XX
CC The present invention relates to nucleic acid sequences
CC (AB067188-AB0671212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 684707 BP; 213423 A; 126375 C; 133438 G; 211468 T; 3 other;

Alignment Scores:
Pred. No.: 9,23e-19 Length: 684707
Score: 318.00 Matches: 123
Percent Similarity: 38.12% Conservative: 76
Best Local Similarity: 23.56% Mismatches: 213
Query Match: 11.72% Indels: 110
DB: 24 Gaps: 17

US-10-066-551-4 (1-525) x AB067196 (1-684707)
QY 4 AsnProLysThrGlnAlaMetProSerGluThrIleSerMetLysThrArgSerLeu 23
DB 258634 AATCCAGCACTACAAAATTC-----AGTACACTCAAGTATGAGAAACCAATTA 258581
QY 24 IleSerLeuLeuCysLeuLeuLeuCysSerCysSerSerTyrProLeuGluGln 43
DB 258580 ATTATGCGATTCATGCAATCCAAACACACCAATGAAATTAACAGCGCTTTCAA 258521
QY 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
DB 258520 AGAATT-----GCCCATTTAACTCCATAGACCGCATTAAG--GGAAATTAATTAAGAA 258470
QY 64 IleArgHisThrProHisThrAsnGlyLeuSerSpleTyrIleLeuAsnAspProHis 83
DB 258469 ATATTA-----ACAAATGT-----GAG 258452
QY 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGln 103
DB 258451 GAACCTTTCCAGTGCCTCCTAGACGCGCTTAGAAGCTGAAACCAATATTCATTCAA 258392
QY 104 TyrTyrIleTyrPargAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAla 123
DB 258391 TATTACATTTTAAACGAGATGCGATTTTACACAGATTCGTGATTTTGGTGGAATAA 258332
QY 124 AlaGluArgGlyValArgValArgLeuLeuLeuAspAspAsnAsnThrArgGlyLeuAsp 143
DB 258331 GCGAAATCCGTTGTGAGTTAGTTATTTGATTTGCTTGATCAACCAACTGAGT 258272
QY 144 AspLeuLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163
DB 258271 AAAGCTTTTACCTCTTGGAAGAAGACGTGAGTATGATTCAGCATTTTGACCA--- 258215

```



```

Db 11086 AGAATT-----GCACATTACGTCACATAGCCGATTAG---GGAATAAATAATTGAA 11136
Oy 64 ILeargHISThrProHISThrAsnGLYLeuSerAspIleTYrLeuLeuAsnAspProHIS 83
Db 11137 ATATTA-----ACAAATGCT-----GAG 11154
Oy 84 GluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHISerLeuAspLeuGln 103
Db 11155 GAAACTTTTCCAGTCCCTCTAGACCGCTTAGAAAGCGTAAACCAATATTCACATTCOA 11214
Oy 104 TYrTYrIleTParArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTYrLeuAla 123
Db 11215 TATTACATTTTAAACGAGATCCGATTTCTACAGAGATTCGATGATTTGGTGAAAAA 11274
Oy 124 AlaGluArgGlyValArgValArgLeuLeuAspAsnAspAsnThrArgGlyLeuAsp 143
Db 11275 GCGAATCCGCTGTTGAGCTTAGGTTATGTTGACGCTTGGATCAAGCAAACTGAGT 11334
Oy 144 AspleuLeuLeuAlaLeuAspSerHISProAsnIleGluValArgLeuPheAsnProPhe 163
Db 11335 AAAGCCTTTTACGCTCTTGAAGAACGCTGAGTTAGTATTCACGCAATTGACCCA--- 11391
Oy 164 ValLeuArgGlySTrParArgAlaLeuGlyTYrLeuThrAspPheProArgLeuAsnArg 183
Db 11392 ATTGCCCTCCCTTGAGTTGA-----AGAACACGCCAATTGAGA 11430
Oy 184 MethISAsnLYSerPheThrAlaAspAsnArgAlaThrIleLeuGlyArgAsnIle 203
Db 11431 AATCACCGTAAATGTTGTGATTCAGCAGCAAAATGCTTACAGTGAGCTTAACATT 11490
Oy 204 GLYAspGLUTYr-----PheLYSValGlyGluAspThrValPheAla 217
Db 11491 GGGGAGAGATATCCCTCTAATACACAGATTTTCTGCTGCGCGATGATACA----- 11541
Oy 218 AspleuAspIleLeuAlaThrGlySerValValGlyGluValSerHISAspPhe----- 235
Db 11542 -----CACATGAAATACACAGCAGCAGCTGTATCGAACTTCAAGATCTTTCTAAT 11595
Oy 236 AspArgTYrTParAlaSerHISerAlaHISAsnAlaThrArgIleIleArgSerGlyAsn 255
Db 11596 GATTGAGTATTATGAAATAATCAGCGCTGCGCTGATCAATTATTAAGAACCTGCA 11655
Oy 256 IleGlyLYSGLYLeuGlnAlaLeuGlyTYrAsnAspGLUTrSerArgHISAlaLeuLeu 275
Db 11656 TTACACAAATATTTTCCACCGATTGATGTAGCTGATGAAAGGCC----- 11700
Oy 276 ArgTYrArgGluThrValGluGlnSerProLeuTYrGlnLYSleGlnThrGlyArgIle 295
Db 11701 -----CAAGTGATATACGGCGGCACATAT 11724
Oy 296 AspTPRGLNserValGlnThrArgLeuIleSerAspAspProAlaLYSGLYLeuAspArg 315
Db 11725 GAT----- 11727
Oy 316 AspArgArgLYSProProIleAlaGlyArgLeuGlnAspAlaLeuLYSGLNProGlnLYS 335
Db 11728 -----AAAGAAAAATGGGCTGCGATTCGATGCTGTGATTGATTGATTCGCTAAAGAA 11781
Oy 336 SerValTYrLeuValSerProTYrPheValProThrLYSserGlyThrAspAlaLeuAla 355
Db 11782 TCAGTTTGATGTATGCGCCCTATTTTCTTCACATGAGAGCGCTTGGCGGTATTCGC 11841
Oy 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375
Db 11842 CCGGCTTGCAATGAGCTGCGTACGAGTACGCTTATATTCCTGCA-----AAAGCGCAT 11895
Oy 376 ValAlaAlaValHISserGlyTYrValLYSArgLYSProLeuLeuLYSAlaGlyIle 395
Db 11896 CCGCGGATTTGATTCACGAGAGTAACTTATGTAAGATGATTAAGACGAGCGCG 11955
Oy 396 LysLeuTYrGluLeuGlnProAsnHISAlaValProAlaThrLYSAspLYSGLYLeuThr 415

```

```

Db 11956 AAAATGATGCTTATCCGATGACTCTTTTGT----- 11988
Oy 416 GlySerSerValThrSerLeuHISAlaLYSThrPheIleValAspGlyArgIlePhe 435
Db 11989 -----CATGCAAAACCATGCTTTGTGGACGAGACGTCGGGCT 12027
Oy 436 IleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValVal 455
Db 12028 ATTGCACTGGCAATTTTGACGTGGCAAGTTTGATGATTCATGATTAATGATGATTC 12087
Oy 456 IleGlySerProLYSleAlaGluIleMetGluArgIleThrLeuAlaAspThrSerProGlu 475
Db 12088 TTATATGATGAA-----AGTGAGCGTATGCTTCATTTAAACGATTTTAAAGAAAT 12141
Oy 476 TYrAlaTYrArgValThrLeuAspArgHISAsnArgLeuGlnTrpHISAspProAlaThr 495
Db 12142 TTT-----GAAGATAGCCGACATA-----TTTACG 12165
Oy 496 ArgLYSThrTYrProAsnGluProGluAlaLYSLeuTrpLYSArgIleAlaAlaLYSle 515
Db 12166 ATGAAAGATATGCAAAATTAACCATTTATGACTGCTATTAAGAAAGTTCTATCCAGTTTA 12225
Oy 516 LeuSer 517
Db 12226 CTATCA 12231

RESULT 15
ID ABA03041 standard; DNA; 2944528 BP.
AC ABA03041:
XX 05-FEB-2002 (first entry)
DT
DT
XX 05-FEB-2002 (first entry)
DE
XX Listeria monocytogenes EGD-e genome sequence.
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW Vitamin B12; bacterial infection; disease; ds.
XX
XX Listeria monocytogenes.
OS
OS
PN WC0200177335-A2.
XX
XX 18-OCT-2001.
PD
XX
XX 11-APR-2001; 2001WO-FR01118.
PF
XX 11-APR-2000; 2000FR-0004629.
PR
XX (INSP ) INST PASTEUR.
PA
XX
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dusurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX Chakraborty T, Domann E, Hahn T, Berche P, Charbit A, Durant L;
XX Garcia-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX Maqueno E, De Pablos B, Wehland J, Kaerst U, Ertlan K, Hauf J;
XX Rose M, Voss H;
XX
XX WPI: 2002-010914/01.
XX
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides
XX
XX Claim 1; SEQ ID No 1; 192pp; French.
XX
XX The present sequence is the genome sequence of Listeria monocytogenes
XX EGD-e. This sequence and fragments of this sequence are useful for
XX selecting probes and primers for detecting genes in L. monocytogenes and
XX related organisms, and to study genetic polymorphisms and other genomes.
XX Proteins (AB47297-AB50149) expressed from the present sequence are

```







KM Infection; ds.  
 XX Listeria innocua.  
 XX W0200228891-A2.  
 XX  
 XX 11-APR-2002.  
 XX  
 XX 04-OCT-2001: 2001MO-FR03061.  
 XX  
 XX 04-OCT-2000: 2000FR-0012697.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Kunst F, Glaser P;  
 XX  
 XX WPI; 2002-332479/37.  
 XX  
 XX New genomic sequences from *Listeria* species, useful for detection,  
 PT treatment and prevention of infection, also related polypeptides,  
 PT antibodies and modulators -  
 XX  
 XX Claim 5; SEQ ID 2058; 180bp; French.  
 XX  
 CC The present invention relates to nucleic acid sequences  
 CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes  
 CC and primers for identification and/or detection of *Listeria* (e.g. as  
 CC contaminants in foods, or mutational analysis) and for analysis of  
 CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 CC used to screen for compounds that modulate gene expression, replication  
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for  
 CC treating infections by *Listeria*, and are useful as immunogens in  
 CC anti-*Listeria* vaccines.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other:  
 Alignment Scores:  
 Pred. No: 2,72e-15 Length: 3011208  
 Score: 293.50 Matches: 113  
 Percent Similarity: 37.61% Conservative: 60  
 Best Local Similarity: 24.57% Mismatches: 172  
 Query Match: 10.82% Indels: 115  
 DB: 24 Gaps: 16  
 US-10-066-551-4 (1-525) x AB069245 (1-3011208)  
 QY 74 SeraspileryleuAsnspProhiscgluAlaPhealaAlaArgAlaLeuile 93  
 Db 2664465 AATGAGATAGAGCTTTTATTTGACGTCACGAAAAATTTATGCGATATT 2664406  
 QY 94 GluSerAlaGluHisSerleuAspLeuGlnTyrTyrIleTyrPArgAsnspIleSercly 113  
 Db 2664405 GAAAAAGCAAAAGATCATATTCATTCATCTACTATTTTCCATTCGATGAGCTTGA 2664346  
 QY 114 ArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeu 133  
 Db 2664345 AACCGTTTGTGCGAGCTTGAAGAAAGCAGACAGAGCTTTAAATGTAATAATTATT 2664286  
 QY 134 LeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisPro 153  
 Db 2664285 TATGATGCAATGGGTTGACGACGACGAAAAAATCATTTTCCGTACTTTCGAAAAAAT 2664226  
 QY 154 AsnIleGluValArgLeuPheAsnProPheValLeuArgGlyTyrPArgAlaLeuGlyTyr 173  
 Db 2664225 GGGCGTTTACTAGACCATTTTCCCATCTAAATTACCA----- 2664187  
 QY 174 LeuThrAspPheProArgLeuAsnArgArgGlyMetHisAsnLysSerPheThrAlaAspAsn 193  
 Db 2664186 TTATATTAATTTC---CGTTTAAATTATATCATCCGAAAAAATTTGGCGATTATTGACGGA 2664130

QY 194 ArgAlaThrIleLeuGlyArgAsnIleGlyAspGluTyrPheLeuValGlyGluAsp 213  
 Db 2664129 GATGTTGGTATATATGTTGGCTTTAATATTTGGACCAAAAT-----CTTGACGCTTCG 2664076  
 QY 214 ThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHis 233  
 Db 2664075 AAAAAATTTCGT----- 2664064  
 QY 234 AspPheAspArgTyrTyrPheAlaSerHisSerAlaHisAsnAlaThrArgIleLeuArgSer 253  
 Db 2664063 -----TATTCGCGCGAT-----ACACATTTTACGCGTCAT 2664034  
 QY 254 GlyAsnIleGlyLysGlyLeuGluAla-----LeuGlyTyrAspAspLeuThrSer 270  
 Db 2664033 GGTAAAGCGGTATATGCTATGACCAACCGCATTTATCATGACGTGAATTCGGCATCATCT 2663974  
 QY 271 ArgHisAlaLeu-----LeuArgTyrArgGluThrVal-----GluGlnSerPro 285  
 Db 2663973 ACTCATTAATAATGACTTAAGACGAGATATTTCCACATTTCCAGGTAAAGCCATACA 2663914  
 QY 286 LeuTyrGlnLysIleGlnThrGly---ArgIleAspTrpGlnSerValGlnThrArgLeu 304  
 Db 2663913 AGCATGCAAAATGTGTCTAGTGTCTGATTCGCAATGCGAACAATAAGAAATGCTTAT 2663854  
 QY 305 IleSerAspAspProAlaLeuGlyLeuAspAspArgAspArgGlyProProIleAlaGly 324  
 Db 2663853 ATTAAA----- 2663848  
 QY 325 ArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPhe 344  
 Db 2663847 -----ATGATAAATGCGCTAAATAAAGATTTACTTGACGTCCTCATTTT 2663800  
 QY 345 ValProThrLysSerclyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364  
 Db 2663799 ATTCCAGATCGACGCTTCTTGAAGCGATAAATTCACACACTTTCAGAGCTGAGTGT 2663740  
 QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaHisSerclyTyrVal 384  
 Db 2663739 CGTGTCAATGATCCGAAT-----AACCTGACCATGCTTTTGTTCACGCAACAAC 2663686  
 QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlyLeuGlnProAsnHis 404  
 Db 2663685 AATTACCGCCGCGTAATGAGAAACAGACGCTAAATCTTC----- 2663644  
 QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424  
 Db 2663643 -----ATTATGATTAATGCTTTT-----ATTCATGCA 2663617  
 QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444  
 Db 2663616 AAGACCTTACTGTTGTGATGCGGAATTCCTTCGTAGAACCTGCATATGACCTTTAGG 2663557  
 QY 445 SerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluGln 464  
 Db 2663556 AGTTTCGCTGAATTTGAATGAAGTCATGCTTTTATATGAAAAAATGCTTCAAAA 2663497  
 QY 465 MetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArg 484  
 Db 2663496 TTAGAGATGCATTTTAGAGAT----- 2663473  
 QY 485 HisAsnArgLeuGlnThrHisAspProAlaThrArgLysThrTyrProAsnGluProGlu 504  
 Db 2663472 -----ATTTGAATGTGATCCATTAACCTGAA 2663443  
 QY 505 -----AlaLysLeuThr-----LysArgIleAlaAlaLysIleLeuSer 517  
 Db 2663442 TTATATGCAAAACGATCTTTATGATTAAATTTAAGAGCGTAAGTCGACTTTATACG 2663383  
 RESULT 18  
 AB070540/c  
 ID AB070540 standard; DNA; 1544 BP.  
 XX

AC ABQ70540;  
XX 29-AUG-2002 (first entry)  
XX  
DE Listeria monocytogenes 4b contig DNA sequence #482.  
XX  
KW Antibacterial; Listeria; food contamination; mutational analysis;  
KM Infection; ds.  
OS Listeria monocytogenes 4b.  
XX  
PN WO200228891-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 04-OCT-2001; 2001WO-FR03061.  
XX  
PR 04-OCT-2000; 2000FR-0012697.  
XX  
PA (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Kunst F, Glaser P;  
XX  
DR MPI; 2002-332479/37.  
XX  
PT New genomic sequences from Listeria species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators  
XX  
PS Claim 14; SEQ ID 3353; 180pp; French.  
XX  
CC The present invention relates to nucleic acid sequences  
CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes  
CC and primers for identification and/or detection of Listeria (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of  
CC gene expression. Proteins encoded by the nucleic acid sequences can be  
CC used to screen for compounds that modulate gene expression, replication  
CC and pathogenicity of Listeria (potential therapeutic agents), also for  
CC treating infections by Listeria, and are useful as immunogens in  
CC anti-Listeria vaccines.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
SQ Sequence 1544 BP; 492 A; 313 C; 240 G; 499 T; 0 other;  
XX  
Alignment Scores:  
pred. No.: 1,3e-19 Length: 1544  
Score: 286.50 Matches: 112  
Percent Similarity: 37.61% Conservative: 61  
Best Local Similarity: 24.35% Mismatches: 172  
Query Match: 10.63% Indels: 115  
DB: 24 Gaps: 16  
US-10-066-551-4 (1-525) x ABQ70540 (1-1544)  
QY 74 SerAspIleuLeuLeuAsnProHisGluAlaPheAlaAlaAlaLeuIle 93  
DB 1274 AATGAGTGAACCTTTTGTAGATGACATGACAAATTCGATTCGCGATATT 1215  
QY 94 GluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleTrpArgAsnAPrIleSergly 113  
DB 1214 GAAAGAGCGAAGATCATATACATCTCATATTATTTTTCATCTGTGAACTGAGC 1155  
QY 114 ArgLeuPheAsnLeuValIleTyrLeuAlaIleGluArgGlyValAlaArgLeuLeu 133  
DB 1154 AATCGCTTAATGGCGGTGGGAAAGAAAGACGACAGAGGTTTAAATGTGAAATATTAT 1095  
QY 134 LeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisPro 153  
DB 1094 TAGATGATCATGGTTCAAGACAGAAAAAATCATTTTCCGTACGTTTCAAAAAAAT 1035

QY 154 AsnIleGluValArgLeuPheAsnProPheValLeuArgIleTrpArgAlaLeuGlyTyr 173  
DB 1034 GGTGGTCTAGTAGACCTTTTTCGCTAAATTACCA----- 996  
QY 174 LeuThrAspPheProAlaGluAsnArgArgMetHisAsnIleSerpPheThrAlaAspAsn 193  
DB 995 TTGATTAACTTT--CCTCAACTACCGAATATCAGCGGAAATGGCTATTAATGATGGA 939  
QY 194 ArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGluTyrPheLeuValGlyGluAsp 213  
DB 938 GATGTTGGTTATATGTGGTGTTTTAATATGTGATCAATAT-----CTGGGGCATCG 885  
QY 214 ThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHis 233  
DB 884 AAGAATTTTGGT----- 873  
QY 234 AspPheAspArgTyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSer 253  
DB 872 -----TACTGGCGTGAT-----ACACATTTACGTGATCAT 843  
QY 254 GlyAsnIleGlyLeuGlyLeuGlnAla-----LeuGlyTyrAsnAspGluThrSer 270  
DB 842 GGAAGAGCGTGTATATGCGATGCAAACTCGCTTATATGATGACTGGAATTCGATCATCC 783  
QY 271 ArgHisAlaLeu-----LeuArgTyrArgGluThrVal-----GluGlnSerPro 285  
DB 782 ACACATATAAATTATTATACAAAGCGCGACTTCCGCTTTCATGCGAAAGGTACATACA 723  
QY 286 LeuTyrGlnIleGlyIleGlnThrGly---ArgIleAspTrpGlnSerValGlnThrArgLeu 304  
DB 722 AGTATGAAATTTGTTTGGAGTGGTCCGATTCGGAATGGCAACAAATCAAAAATGGTTAC 663  
QY 305 IleSerAspAspProAlaIleGlyLeuAspArgAspArgIleProIleAlaGly 324  
DB 662 ATCAAA----- 657  
QY 325 ArgLeuGlnAspAlaLeuIleGlyGlnProGlyIleSerValTyrLeuValSerProTyrPhe 344  
DB 656 -----ATGATCAATGCTGCTTAAAAAACAACTTTATACAACTCCCTATTAT 609  
QY 345 ValProThrIleSerpGlyThrAspAlaLeuAlaIleLeuValGlnAspGlyIleAspVal 364  
DB 608 ATTCCAGATCGAAGTTTGTGTAAGCAATCAAAATCGCACACTTCTGGGGTAGATGTC 549  
QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerglyTyrVal 384  
DB 548 AGGTTATGATTCCTAAT-----AAGCCTGACCAATGCTTTGTTTACCGCGAATCACT 495  
QY 385 LysTyrArgIleSerpLeuLeuIleGlyIleLeuLeuTyrGluLeuGlnProAsnHis 404  
DB 494 AACTACCCCGGTGAATTAATGGAACCGGAGCGAATAATCTTT----- 453  
QY 405 AlaValProAlaThrIleAspIleGlyLeuThrGlySerSerValThrSerLeuHisAla 424  
DB 452 -----ATTATGATTAATGGTTT-----ATCCACGCG 426  
QY 425 LysThrPheIleValAspGlyIleSerpIlePheIleGlySerPheAsnLeuAspProArg 444  
DB 425 AAAACACTGCTGTAGATGCGAATATGCTTCAAGTTGGAACCGCAATATGATATTCGCG 366  
QY 445 SerIleAlaArgLeuAsnThrGluMetGlyValIleIleGluSerProIleAlaGluGln 464  
DB 365 AGTTTCGACTTAATTTTGAAGTAATGCTTTATATTATGAAAAACAAAGGTGCAAAA 306  
QY 465 MetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArg 484  
DB 305 CTAGAACATCCCTTTTGGAGAT----- 282  
QY 485 HisAsnArgLeuGlnThrHisAspProAlaThrArgIleSerpProAsnGluProGlu 504  
DB 281 -----ATCTAAATATCGATACATCAACACCTGAA 252  
QY 505 -----AlaIleLeuTyr-----LysArgIleAlaIleAlaIleLeuSer 517

DB 251 TTATATGCAAAAGACTTATGATTAATTAAGACGGGTAAGCCGACTTTATGCG 192  
 RESULT 19  
 ID ABA03041/C  
 ID ABA03041 standard; DNA: 2944528 BP.  
 XX ABA03041;  
 AC ABA03041;  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes EGD-e genome sequence.  
 XX  
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease; ds.  
 XX  
 OS Listeria monocytogenes.  
 XX  
 PN W0200177335-A2.  
 PD 18-OCT-2001.  
 XX  
 PF 11-APR-2001; 2001WO-FR01118.  
 XX  
 PR 11-APR-2000; 2000FR-0004629.  
 XX  
 PA (INSP) INST PASTEUR.  
 XX  
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Eshti H, Dehoux P,  
 PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P,  
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,  
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
 PI Chakraborty T, Dommann E, Hain T, Berche P, Charbalt A, Durant L,  
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
 PI Medunovic E, De Pablos B, Wehland J, Kaerst U, Ertian K, Hauf J,  
 PI Rose M, Voss H;  
 PI  
 DR WPI; 2002-010914/01.  
 XX  
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
 PT and prevention of Listeria and related bacterial infections, and  
 PT related polypeptides  
 XX  
 PS Claim 1; SEQ ID No 1; 192pp; French.  
 XX  
 CC The present sequence is the genome sequence of Listeria monocytogenes  
 CC EGD-e. This sequence and fragments of this sequence are useful for  
 CC selecting probes and primers for detecting genes in L. monocytogenes and  
 CC related organisms, and to study genetic polymorphisms and other genomes.  
 CC Proteins (ABBA4297-ABBS0149) expressed from the present sequence are  
 CC useful for raising specific antibodies, identification of L.  
 CC monocytogenes and related organisms, and for biosynthesis and  
 CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
 CC proteins encoded by it are also useful for selecting compounds that  
 CC regulate gene expression and cell replication and modulate L.  
 CC monocytogenes-related diseases. In addition, this sequence and proteins  
 CC encoded by it are useful in pharmaceutical and vaccine compositions for  
 CC the treatment or prevention of infections by L. monocytogenes and related  
 CC organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 8.62e-15 Length: 2944528  
 Score: 288.50 Matches: 112  
 Percent Similarity: 37.61% Conservative: 61  
 Best Local Similarity: 24.35% Mismatches: 172  
 Query Match: 10.63% Indels: 115  
 DB: 24 Gaps: 16

US-10-066-551-4 (1-525) x ABA03041 (1-2944528)  
 QY 74 SeraspileryleuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIle 93  
 DB 2579993 AATGAAGTGAACCTTTTGTGACGCGACAGAGAAATTCGATCGCTAATGCTGATAT 2579994  
 QY 94 GluSerIaGluHisSerLeuAspLeuGlnTyrTyrIleTyrPArgAsnAspIleSerGly 113  
 DB 2579933 GAAAGACGAAAGATCATATACATCTCATATATATTTTTCATCCGATGACATAGCC 2579984  
 QY 114 ArgLeuLeuPheAsnLeuValTyrLeuAlaIaGluArgGlyValArgValArgLeuLeu 133  
 DB 2579873 AATCGCTTAATGCGCGCTCGAAGAAAGACGACAGCAAGTTAATGTGAAATATAT 2579814  
 QY 134 LeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisPro 153  
 DB 2579813 TATGATGCAATGGGATCAAGACAGCAAGAAATCATTTTCCGACCTTCAGAAATAT 2579754  
 QY 154 AsnIleGluValArgLeuPheAsnProPheValLeuArgTyrPArgAlaLeuGlyTyr 173  
 DB 2579753 GCGCGTTTACTCAGACCATTTTTCATCTAATATACCA----- 2579715  
 QY 174 LeuThrAspPheProArgLeuAsnArgArgGlyHisAsnLysSerPheThrAlaAspAsn 193  
 DB 2579714 TTGATTAATTTTCCGCTCACTACCGAATCATGAAATTAATGCTAATATGATGGA 2579658  
 QY 194 ArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlyTyrPheLysValGlyGluAsp 213  
 DB 2579657 GACGTAGCTTAATATGATGCTTTTAATATGATGATCAATAT-----CTTGGGCGCATCG 2579604  
 QY 214 ThrAlaPheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyValSerHis 233  
 DB 2579603 AATAAATTTGCT----- 2579592  
 QY 234 AspPheAspArgTyrTyrPArgAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSer 253  
 DB 2579591 -----TACTGCGCTGAT-----ACACATTTGCGCTGAT 2579562  
 QY 254 GlyAsnIleGlyLysGlyLeuGlnAla-----LeuGlyTyrAsnAspGluThrSer 270  
 DB 2579561 GGAAGACCTTTATGCGATGCAACATCGCTTTATATGATGATGGAATTTGCTGATCGTCC 2579502  
 QY 271 ArgHisAlaLeu-----LeuArgTyrArgGluThrVal-----GluGlnSerPro 285  
 DB 2579501 ACACATTAATGATTAACAAGCGGCTACTCCGACTTTTCATGGGAAGGTCATACA 2579442  
 QY 286 LeuTyrGlnLysIleGlnThrGly-----ArgIleAspTyrGlnSerValGlnThrArgLeu 304  
 DB 2579441 AGTATGCAATTTGTTGAGTGGCCCTGATTCGGAATGCAACAAATCAAAATGCTTAC 2579382  
 QY 305 IleSerAspAspProAlaLysGlyLeuAspArgAspArgLysProProlleAlaGly 324  
 DB 2579381 ATCAAA----- 2579376  
 QY 325 ArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValTyrLeuValSerProTyrPhe 344  
 DB 2579375 -----ATGATCAATGCGCGCAAAACAAATTAATTAACATCACCTTATTTT 2579328  
 QY 345 ValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364  
 DB 2579327 ATTCGATGCAATGTTTACTTCTGAAATCAATCCAGACACTTTTGGGATGATGATTC 2579268  
 QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaHisSerGlyTyrVal 384  
 DB 2579267 CGTGTATGATTCCTAAT-----AGCCAGACCATCTTTTGTATACCGCGCAACAAT 2579214  
 QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlyLeuGlnProAsnHis 404  
 DB 2579213 AATTATCCCGTGAATTAAGAAACAGGTGCAAAATCTTT----- 2579172  
 QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424  
 DB 2579171 -----ATTATGATTAATGCTTT-----ATTCACGCA 2579145

QY 425 LysThrPheIleValaspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444  
DB 2579144 AAAACCCCTGCTTAGATGATGTAATGCTTACGATGGAACGCAATATGATTCCTG 2579085  
QY 445 SerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluIn 464  
DB 2579084 ACCTTCCGATTAATTTAAAGTAATGCTTTATTATGAAAAACAAATGCTGCAAAA 2579025  
QY 465 MetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArg 484  
DB 2579024 CTAGAAAGATGCTTTTTCGAACAT 2579001  
QY 485 HisAsnArgLeuGlnIntPrHisAspProAlaThrArgLysThrTyrProAsnGluProGlu 504  
DB 2579000 2579000 -ATTTAAATCGATCACTACACCTGAA 2578971  
QY 505 -AlaLysLeuTyr-----LysArgIleAlaAlaLysIleLeuSer 517  
DB 2578970 TTATATGCAAAACGATCTTATGATTAATTTAAAGAACGCGTAAAGCGACTTTATCG 2578911

RESULT 20  
ABK73787  
ID ABK73787 standard; DNA; 1446 BP.  
AC ABK73787;  
XX  
XX 13-AUG-2002 (first entry)  
DT  
DE Bacillus licheniformis genomic sequence tag (GST) #1078.  
KM Differential gene expression: genomic sequenced tag; GST;  
KW altered culture condition: environmental stress;  
XX physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
PF 05-OCT-2001; 2001WO-US31437.  
PR 06-OCT-2000; 2000US-0680598.  
PR 27-MAR-2001; 2001US-279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
XX WPI; 2002-416684/44.  
DR  
PT Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second  
PT Bacillus cells, by using substrate containing Bacillus genomic  
PT sequenced tag array  
XX  
PS Claim 4; SEQ ID NO 1078; 2000P; English.  
XX  
XX The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way

CC in which Bacillus cells adapt to changes in culture conditions,  
CC environmental stress or other physiological provocation. Extensive  
CC follow-up characterisation is unnecessary, when one spot on an array  
CC equals one gene or one open reading frame, since sequence information is  
CC available. This sequence represents a genomic sequence tag (GST) used in  
CC the method of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1446 BP; 411 A; 325 C; 321 G; 389 T; 0 other;  
XX  
Alignment Scores:  
Pred. No.: 7,886-19 Length: 1446  
Score: 280.50 Matches: 110  
Percent Similarity: 39.59 Conservative: 65  
Best Local Similarity: 24.89 Mismatches: 152  
Query Match: 10.34 Indels: 115  
DB: Gaps: 16

US-10-066-551-4 (1-525) x ABK73787 (1-1446)  
QY 76 IleTyrLeuLeuAsnAspProHisGluAlaPheAlaIleArgAlaIleGluSer 95  
DB 360 GTCGAGCTGTTAAACGCGAGCGCAAGAAATTCACAGCTGCTCGCATTTGAAAC 419  
QY 96 AlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrArgAsnAspIleSerGlyArgLeu 115  
DB 420 GCCAAGACCATATTCATCTTCATATATTTATTAAGCCATGATGCGCAAAAG 479  
QY 116 LeuPheAsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAsp 135  
DB 480 CTCGGGATCGCTCATCAAAAAGCAAAAGAGCGTTGAAGTCCGCTTTGTATGAT 539  
QY 136 AspaAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155  
DB 540 GAATTAAGATCAAGAAAGCCTCAGG----- 563  
QY 156 GluValArgLeuPheAsnProPheValIleuArgLysTyrArgAlaLeuGlyTyrLeu 174  
DB 564 -----AAAACCTTTTCAAAAGACCTGAAAAAACCGGAGGATTTGCGAA 608  
QY 175 ThrAspPhePro-----ArgLeuAsnArgArgMetHisAsn 186  
DB 609 GTCTTTTCCCGTCGCGCTTAATTTATTCATATTCGATCAATTAAGCAACACCGC 668  
QY 187 LysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlu 206  
DB 669 AAGCTGCTCATTTATCGAGCGGCTCATCGCTATGTCGAGGATTTAATGCGAGACGAA 728  
QY 207 TyrPheLysValGlyGluAspThrValPheAla-----AspLeuAspIleLeuAla 223  
DB 729 TAT-----CTCGCTCGAACCAGCGCTTGGCTATTGGCGGACACGCACTCAGAAATC 782  
QY 224 ThrGlySerValAlaGlyGluValSerHisAspPheAspArgTyrTrp-----AlaSer 241  
DB 783 AAAGCAACAGCGCTCCATCGATTCAAACGCTTTATCCTGGATGCAATCAGGCTCG 842  
QY 242 HisSerAlaHisAsnAlaThr-----ArgIleIleArgSerGly 254  
DB 843 CACCAT--CACGACATTAATATATATGCAAAACCATTTTCTGACATCGATCCACGA 899  
QY 255 AsnIleGlyLysIleGluGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 274  
DB 900 AACATGGCATGCAAAATCGCACACGCGCCGATTCAGAA----- 941  
QY 275 LeuArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrArgArg 294  
DB 942 -----TGGAGCAGATTTAAAGGCTAT 965  
QY 295 IleAspTyrGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAsp 314  
DB 111

```

Db 966 ATT-----AAATGATATCA-----980
QY 315 ArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlu 334
Db 981 -----ACGGCAAG 989
QY 335 LysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeu 354
Db 990 CGGTCATCTTATATACAGCCCTATTATACCGAGCAACCGCTCGATCGCCCTC 1049
QY 355 AlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThr 374
Db 1050 AGGATCCCTGTCTGTCCGGCTTGACGTACGATCATGATTCCTTAC-----AAACCC 1103
QY 375 AspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGly 394
Db 1104 GACCATCCGTTGTGTCTGTCGGCAGCACTATTCTTACATCGCGCACTTTTAAAGCAGGC 1163
QY 395 IleLysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeu 414
Db 1164 GCCTCCGTAATT-----ATTACGACCAAGGTTT 1193
QY 415 ThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIle 434
Db 1194 -----ATCCATGCGAAGAACGATTGAGTAGAGAGAAATTTCA 1232
QY 435 PheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyVal 454
Db 1233 TCGGTCGGCAGCGCAATATTGATGATCGCAGCTTCAAGCTCAATTTTGAAGTAAATGCA 1292
QY 455 ValIleGluSerProLysIleAlaGluGlnMet-----GluArgThrLeuAla 470
Db 1293 TTTTATATGATGAGAAATGCGAAGACCTGTCTCATCATTTGAAAAAGATCTG-- 1349
QY 471 AspheSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThr 490
Db 1350 CAAGTGTCAAGGAGCTGACCTTTGAGCAATTTTAAACAGAACCAATCGATCCGCTTC 1409
QY 491 HisAsp 492
Db 1410 AAGAA 1415

RESULT 21
AAV24658
ID AAV24658 standard; DNA; 618 BP.
XX
AC AAV24658;
XX
DT 26-JUN-1998 (first entry)
XX
DE H. pylori ORF 05cp1191lorf35.
XX
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
KW bacterium; ds.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT CDS 1..618
FT /*tag= a
FT /note= "no stop codon given"
XX
PN W09737044-A1.
XX
PD 09-OCT-1997.
XX
PF 27-MAR-1997; 97MO-US05223.
XX
PR 06-DEC-1996; 96US-0761318.
PR 29-MAR-1996; 96US-0625811.
PR 02-APR-1996; 96US-0758731.

```

```

PR 25-OCT-1996; 96US-0736905.
PR 28-OCT-1996; 96US-0738859.
XX
PA (ASTR ) ASTRA AB.
XX
PI Alm RA, Smith D;
XX
DR WPI: 1997-503122/46.
DR P-PSDB; AAW55249.
XX
PT Helicobacter pylori nucleic acid sequences and encoded
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
PT infection and for diagnosis of H. pylori infection
XX
PS Claims 5, 6; Page 167; 1145pp; English.
XX
CC This sequence encodes a Helicobacter pylori protein of unspecified
CC function.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds,
CC useful as potential H. pylori life cycle activators or inhibitors.
CC The DNA and probes derived from it may be used for the
CC identification of H. pylori in a sample, and the diagnosis of
CC H. pylori infection. Nucleic acid sequences complementary to the
CC DNA act as antisense sequences, and can be used to prevent the
CC translation of H. pylori mRNA. Antibodies against the protein can
CC be used in immunoassays to evaluate the abundance and distribution
CC of H. pylori-specific antigens. The genomic sequence of H. pylori
CC (ATCC 35679) was determined from overlapping contigs generated
CC by mechanically shearing the bacterial DNA. The sequences were
CC analysed for ORF of at least 180 nucleotides, and the predicted
CC coding regions defined by computer evaluation. To identify likely
CC H. pylori antigens for vaccine development, the amino acid
CC sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having
CC identified and determined the sequences of interest, particular
CC regions can be isolated from H. pylori by PCR amplification for
CC recombinant polypeptide production, e.g. in E. coli hosts.
XX
SQ Sequence 618 BP; 185 A; 124 C; 121 G; 188 T; 0 other.

Alignment Scores:
Pred. No.: 2,86e-19 Length: 618
Score: 279.50 Matches: 66
Percent Similarity: 53.37% Conservative: 45
Best local Similarity: 31.73% Mismatches: 86
Query Match: 10.30% Indels: 11
DB: 18 Gaps: 5

US-10-066-551-4 (1-525) x AAV24658 (1-618)
QY 321 ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuVal 340
Db 16 CCCATCAAAATCGCTTTTGAAGAAAGCCCTTAAAGCGTAAGATCGTTTATATCGCT 75
QY 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuLysValGlnAsp 360
Db 76 TCATCGTATTTTATCCAGTAAAGATATGAAATCTTAAATCAATTTCTAAG 135
QY 361 GlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaHis 380
Db 136 GGGATGATGATGATATCTTACCAATTCCTTCATCTACGACGCTATCGTGTCAT 195
QY 381 SerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeu 400
Db 196 GGGCGGTGGAAAGTATCGCAATTAATTAAGTGGCGGCAATGCTATGAATA 255.
QY 401 GlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThr 420
Db 256 CGAAGCATTTTTCACACCGCAGATTAAGGCGC-----TTTACGACCAACAT 306
QY 421 SerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsn 440

```

Db 307 TCCTACACGGCAGACATGTTTGTGACGATGCCCTTAACGCTTCTAGGAGATTTAAT 366  
 Qy 441 LeuaspProargSerAlaArgLeuasnThrGluMetGlyValIleGluSerProlys 460  
 Db 367 ATTAGTCGGCGCTGATATTAACACTGAAGTGGCTCTGTTGTGACACCCGCTCT 426  
 Qy 461 IleAlaGluGlnMetGluArgThrIleuAlaAspThrSerProGluTyrAlaTyrAla 480  
 Db 427 TTGGCCAAAGGGGTGGCTTTGTGCTTAAGATCATGCC--CAACATCATGGCATTTA 483  
 Qy 481 ThrLeuAspArgHisAsnAlaGluGlnThrPheAspProAlaThrArgLysThrTyr--- 499  
 Db 484 GTCTTGATCGCAT---ACAGTGATTTGGGAA-----GCTACAGAAGAGAGCATCTTA 534  
 Qy 500 -----ProAsnGluProGluAlaLysLeuThrLysArgIleAlaAlaLysIleLeu 516  
 Db 535 ATCCATGAAAAAAGCTGCGTATACCTCTTCTTTGCGTTGATTAAAGATGCTCT 594  
 Qy 517 SerLeuLeuProIleGluSerLeu 524  
 Db 595 AAGTCCCTCTGAAAGAACTT 618

## RESULT 22

ABA51633  
 ID ABA51633 standard: DNA; 1482 BP.

XX ABA51633;

DT 23-JAN-2002 (first entry)

XX Staphylococcus aureus polynucleotide SEQ ID NO 21.

DE Staphylococcus aureus; antimicrobial; growth rate; infection; vaccine;

KW Immunity; ds.

OS Staphylococcus aureus.

XX MO200177365-A2.

XX 18-OCT-2001.

PF 06-APR-2001; 2001MO-US11177.

PR 06-APR-2000; 2000US-266327P.

PA (PHAA) PHARMACIA & UPJOHN CO.

PI Tomlch CC;

PI MPI: 2002-010920/01.

XX P-PSDB; ABB32481.

PS Claim 1; Fig 1; 67pp; English.

The invention relates to identifying antimicrobial agents that target the  
 CC proteins (ABA32471-ABB32494) encoded by genes (ABA51623-ABA51634 and  
 CC ABA51659-ABA51670) essential for the survival of staphylococcus bacteria,  
 CC especially staphylococcus aureus, including antimicrobial agents that  
 CC interfere with the expression of essential coding sequence products and  
 CC antimicrobial agents that interfere with the function of essential coding  
 CC sequence products. The genes and proteins are useful for identifying  
 CC agents that decreases the growth rate of a microbe preferably S. aureus.  
 CC The agents identified by the above method are bactericidal and will be  
 CC useful for treating a subject infected with S. aureus or at risk of being  
 CC infected by S. aureus and may be used in vaccines for producing  
 CC immunological response to S. aureus resulting in immunity. The present  
 CC sequence is that of an essential S. aureus polynucleotide coding

XX SQ Sequence 1482 BP; 521 A; 189 C; 270 G; 502 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 3,016-18 Length: 1482  
 Score: 275.00 Matches: 100  
 Percent Similarity: 38.84% Conservative: 81  
 Best Local Similarity: 21.46% Mismatches: 149  
 Query Match: 10.14% Indels: 136  
 DB: 24 Gaps: 16  
 US-10-066-551-4 (1-525) x ABA51633 (1-1482)  
 Qy 74 SerArgIleTyr-----LeuLeuAsnAspProHisGluAlaPheAlaAlaAlaAla 91  
 Db 415 AATGATTTATATGATCAAGTTTAAAGAT----- 444  
 Qy 92 LeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleThrPheAsnAspIle 111  
 Db 445 --ATTAAATGCAAAAGAAATATATCCATTTAGAGTACTTACTTCCGCTTAGATGCT 501  
 Qy 112 SerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArg 131  
 Db 502 TTAGGTAAAGAAATTTTACATGCTTTAGAGAAAGAAATTTGAACAGCTTAGAGTAA 561  
 Qy 132 LeuLeuLeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer 151  
 Db 562 ATATTATATGATGATGCTGATCTAA----- 588  
 Qy 152 HisProAsnIleGluValaArgLeuPheAsnProPheValaLeuArgLysTyrArgAlaLeu 171  
 Db 589 -----AATGTTAAGATGGCAAAATTTTGATCATTTT-----AATCGTTA 627  
 Qy 172 Gly-----TyrLeuThrAspPhePro-----ArgLeuAsn 181  
 Db 628 GGTGCAAGATTTGACATTTTGGCTGCAAAATTTACCGTATTGATTTGACAAATGAA 667  
 Qy 182 ArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyArg 201  
 Db 688 AATGAATCATGATAAAATCATGATCATGATGATGATGATGATGATGATGATGATGAT 747  
 Qy 202 AsnIleGlyAspArgLysThrPheLysValGlyGluAspThrValPheAlaAspLeuAspIle 221  
 Db 748 AACATTGGTATGATATCTAGGATTTAGAAATTTAGATATTGAGAGATGACCAATTA 807  
 Qy 222 LeuAlaThrGlySerValaGlyGluValaSerHisAspPheAspArgTyrTyrAlaSer 241  
 Db 808 CGTATCAAGGGGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 867  
 Qy 242 HisSerAlaHisAsnAlaThrArgIle-----ArgSerGly 254  
 Db 868 --CAAGCCGACCGTCACAAATTTGAATATGATTTAAGTATTTCCCTAAAGAAAGCGA 924  
 Qy 255 AsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspLysThrSerArgHisAlaLeu 274  
 Db 925 CCATTGGC----- 933  
 Qy 275 LeuArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrArg 294  
 Db 934 -----AATTCACCAATTCAAATAGCTGCAAGTGGCCGGCT 969  
 Qy 295 IleAspTrpGlnSerValGln-----ThrArgLeuIleSerAspAspProAlaLys 311  
 Db 970 AGTACTGGCATCAATTGAAATGCGTTATACAAATGATTT----- 1011  
 Qy 312 GlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLys 331  
 Db 1012 -----Arg 1014  
 Qy 332 GlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThr 351  
 Db 1015 AGTCAAGAAATCTGTATATTTACATACCAATTTTTCATTCGGATTAATTCATATATA 1074



[illegible]

XX	PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	PI	Doucette-Stamm LA, Bush D;
XX	PI	WPI; 2002-381255/41.
XX	DR	P-PSDB; ABP39331.
XX	PT	Novel isolated nucleic acid encoding a <i>Staphylococcus epidermidis</i>
XX	PT	polypeptide, useful for diagnosing and treating bacterial infections -
XX	PS	Disclosure; SEQ ID 1339; 267pp; English.
XX	XX	
CC	CC	ABN90538 to ABN93374 represent <i>Staphylococcus epidermidis</i> open reading
CC	CC	frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC	CC	given in ABP5124 to ABP3960. The <i>S. epidermidis</i> sequences have
CC	CC	antibacterial activity and can be used in gene therapy. The sequences
CC	CC	can also be used in the diagnosis and treatment of bacterial infections,
CC	CC	particularly <i>S. epidermidis</i> infections. The sequences can be used to
CC	CC	screen for compounds able to interfere with the <i>S. epidermidis</i> life
CC	CC	cycle or inhibit <i>S. epidermidis</i> infection.
CC	CC	N.B. The sequence data for this patent did not form part of the printed
CC	CC	specification, but was obtained in electronic format directly from the
CC	CC	USPTO web site.
XX	XX	
XX	XX	Sequence 1485 BP; 523 A; 202 C; 262 G; 498 T; 0 other;
XX	XX	
XX	XX	Alignment Scores:
XX	XX	Pred. NO.: 3.65e-17 Length: 1485
XX	XX	Score: 264.50 Matches: 101
XX	XX	Percent Similarity: 38.05% Conservative: 71
XX	XX	Best Local Similarity: 22.35% Mismatches: 177
XX	XX	Query Match: 9.75% Indels: 103
XX	XX	DB: 24 Gaps: 11
XX	XX	
XX	XX	US-10-066-551-4 (1-525) x ABN91876 (1-1485)
XX	XX	
QY	76	ILEYRLEULEAANAAPRROHISGLUALAHEHALEHALEAALALEULEGUSER 95
DB	337	ATTCGATTTATTTACAGAGGTGCATTAAGCGTTTATATGAAAAGTACTTGAGGATATTTACAT 456
QY	96	ALAGLHHISSEIRLEAASPLEAGLITTYRYYLLETPRAGSNAISPIESEGLYARLEU 115
DB	457	GCTCAAGACTATATACATCTAGTACTACTATACCTTGACCTGATGATTAGGTAAGAAGA 516
QY	116	LEUPHEASPLEUVALTYRLEUALALEGLUARGGLYVALARVALARGLEULEUASP 135
DB	517	ATCTTAGAGTGCACCTTGAAACTTAACCTTAAGAACAGCTTTAGAACCTTTGTATAGAC 576
QY	136	ASPSNAASNTTATATGGYGLYLEUASPRLEUULEUUALALEUSPERHISPROASILE 155
DB	577	GATGTGGTCTTAA----- 591
QY	156	GLUVALARGLEUPHEASNPORPHEVALLEUARGLYSTRPRAGALALEUGLYTYRLEUTHR 175
DB	592	AAGGTACATATATCAAAA-----TTTAAACATTTCCAGACATAGCTGAGAGATT 642
QY	176	ASP-----PHEPCO-----ARGLEAENARAGARGMEHIS 185
DB	643	GAAGCATTTTCCCTTCGAAAGTACCTTTAATCAATTTCCGAAGATGAATATCGAATCAT 702
QY	186	ASNYSSERPHETHPALIASPASNARGALATHRIELEUGLYGLYADASNIIEGLYASP 205
DB	703	AGMAAGATTTATCATTAATAGATGACGACAAATTCGTTACTGCGGTTTAAATGTCGCGGAT 762
QY	206	GLUTYRPHELYVALIGLYGLUASPRTHVALPHEALIASPLEUASPILEUUALATHRGLY 225
DB	763	GATTATTTAGCATTAAGTAAGTATAGTTACTGACAGATGATACATACACCTGTTCACAGT 822
QY	226	SERVALVALIGLYLVALISERHISASPRHEASPRGRGTYRPRALASERHISSERALAHIS 245
DB	823	GAATGACATGCATGACATPACAAATTAAGATTAATTTTAACTGGAATTCACAGTGC----- 876





```

Oy 46 GUserArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGlnIleArg 65
Db 325 -----TATAATACGACGACATTTTAAACAGACAGACAC----- 357
Oy 66 HisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisGluAla 85
Db 358 -----GATTTAARRRRRTATACACACAGCCGACGCAAAAAA 390
Oy 86 PheAlaAlaArgAlaLeuIleGlnSerAlaGluHisSerLeuAspLeuGlnTyrTyr 105
Db 391 TTGATGACCTAATACACACATCCGTAATGCTACGATTAATTAATTCATTTTCAGTACTAT 450
Oy 106 IleTyrPArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleGlu 125
Db 451 ATTATTCAAAATGATGCAATAGTGTCTACCATTTTAATGAACTTGTAATAAAGCGGAA 510
Oy 126 ArgGlyValArgValArgLeuLeuAspAspAsnAspThrArgGlyLeuAspAspLeu 145
Db 511 CAAGGTGTAAGAAGTAAATAATCTTATGATGACATGGGTTCTCGTGACGTG----- 561
Oy 146 LeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPheValLeu 165
Db 562 -----CGTAAAAAGCGCTTACGCCCGCTTTGCGCAT 591
Oy 166 ArgLysTyrPArgAlaLeuGlyTyrLeuThrAspPhePro----- 178
Db 592 AAAGGTGACATGCTGAACCATTT-----TTCCATCAAAATTTACCTTAATTAAC 642
Oy 179 ---ArgLeuAsnArgArgGlyHisAsnLysSerPheThrAlaAspAsnArgAlaThrIle 197
Db 643 TTCGGATATGAACAATCGAATCGAATCGAATGTTGTAATAGTAGGGAATTTGGATAT 702
Oy 198 LeuGlyGlyArgAsnIleGlyAspGlyTyrPheLysValGlyGluAspThrValPheAla 217
Db 703 GTTGGGTGTTTAATCTGTGATGAGTAC-----TTAGCTAAATCAAAAAAATTCGGC 756
Oy 218 -----AspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAsp 234
Db 757 TATTTGGCGAGATACGATTTACGAATTTGCGGAGATGCGATGATGCAATTCAGCA 816
Oy 235 PheAspArgTyrTyrAlaSerHisSerAlaHisAsnAlaThrArg----- 249
Db 817 TTATATTTAGATTGGAATTC-----CAAGCCACACGTCGACCATCTCTCTAT 864
Oy 250 -----IleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAla 262
Db 865 GATGATCGTTATTTCCCATGTGAATTTCTGTGGACACATTTGCGTCAATAGCTTCT 924
Oy 263 LeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGlu 282
Db 925 ACTGGTCTCGACGAGAA----- 942
Oy 283 GlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTyrGlnSerValGlnThr 302
Db 943 -----TGGGACAGATTAATAATACGGCTATATTTG----- 969
Oy 303 ArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysProPheIle 322
Db 970 AAAATGATTTCA----- 981
Oy 323 AlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerPro 342
Db 982 -----TCTGCTAAAAAATCGATTAATTAATTCATCTCC 1014
Oy 343 TyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIle 362
Db 1015 TATTTTCATACCGATCAACCTTTTATGATTTATTAATAATGCGGCATTAGGTGTT 1074
Oy 363 AspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGly 382
Db 1075 GATGTCATATATCATGATTCCTAAT-----AAACCTGACCATCCGTTGTTTGGGCT 1128

```

```

Oy 383 TyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnPro 402
Db 1129 ACTTTAAAAAATGACGACATCTTATTAATGATCCGCGTTAAAGTATTCATCACTAC----- 1182
Oy 403 AsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeu 422
Db 1183 -----GACCAATGGCTTT-----TTA 1197
Oy 423 HisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAsp 442
Db 1198 CACTCAAAACACTTGTATATGATGATGAATTCGAAGTGTGGAACGCTAATATGAC 1257

RESULT 26
AAK13075/C
ID AAK13075 standard; DNA; 28690 BP.
XX
AC AAK13075;
XX
DT 19-MAR-1999 (first entry)
XX
DE Enterococcus faecalis genome contig SEQ ID NO:138.
XX
KW Enterococcus faecalis; config; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
OS Enterococcus faecalis.
XX
PN W09850555-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98W0-US08985.
XX
PR 14-NOV-1997; 97US-0066009.
XX 06-MAY-1997; 97US-0044031.
XX 16-MAY-1997; 97US-0046555.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Dillon PJ, Kunsch CA;
XX
DR WPI: 1999-045171/04.
XX
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
PS Claim 1: Page 816-830; 2084pp; English.
XX
XX
XX A computer readable medium has been developed which has recorded on it
XX 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX AAK12938 to AAK13919 represent these nucleotide sequences which are
XX primary nucleotide sequences, also known as contigs. The computer-based
XX system can identify fragments of the Enterococcus faecalis genome with
XX commercial importance. The products can be used to detect the presence
XX of Enterococcus faecalis in samples. They can also be used for
XX diagnosis of Enterococcal infection in an animal and monitoring
XX progression of disease, and for identifying agents which can be used to
XX modulate the growth or pathogenicity of Enterococcus faecalis, or
XX another related organism, in vivo or in vitro. In particular the
XX polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX can be used in vaccines to prevent or attenuate an Enterococcal
XX infection.
XX
XX
XX Sequence 28690 BP; 8977 A; 5769 C; 4894 G; 9006 T; 44 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1.26e-13 Length: 28690
XX Score: 248.50 Matches: 105
XX Percent Similarity: 38.37% Conservatve: 60
XX Best Local Similarity: 24.42% Mismatches: 184
XX Query Match: 9.16% Indels: 81

```

DB:	20	gaps:	14
US-10-066-551-4 (1-525) x AAX13075 (1-28690)			
OY 40	ProLeuGluGluArgThyGluSerArgHisPheAsnThrSerLysProValArgLeuAsp		59
	::: ::::     ::::		
Db 28537	CCGACGGAGGAGCAACACAGTTTACACACTTTTTCCTCG		28496
OY 60	AsnIleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeu		79
	:::                 ::::   ::: ::::		
Db 28495	-----TTAAATAGCATGCCCTGCATAG-----AAAAATGATGTCAATTTT		28451
OY 80	AsnAspProHisGlyAlaPheAlaIleArgAlaIleLeuIleGlySerAlaGlnHisSer		99
	:::    :::   :::    :::   :::		
Db 28450	ACGGATGGTACCGGAAAAATTCAATCCGTGATGGCGATATRAAAAAAGCTACGACTCC		28391
OY 100	LeuAspLeuGlnTyrTyrIleTrrArgAsnAspIleSerGlyArgLeuLeuPheAsnLeu		119
	::: ::::   :::		
Db 28390	ATTCAATATGAAATTTATATGCTTTGTATACGGATCATATGACACCAAAATTTTGAAC		28331
OY 120	ValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuLeuAspAsnAspThr		139
	:::		
Db 28330	TTACAGCAAAAGGCGGCTGAAGGGGTGAAGTGGCGTTCCTTTACGATCCCTTGCTCA		28271
OY 140	ArgGlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIleGlyValArgLeu		159
	:::		
Db 28270	AAGGC-----ACCAAGTTCATCATTTGAAATGAATGAATAA		28232
OY 160	PheAsnProPheValLeuArgLysTrrArgAlaLeuGlyTyrLeuThrAspPheProArg		179
Db 28231	AACGGTGGTTTGTCCAAACGTTTATATCTCTCAAAAGCACTTTGAAGTT--CGT		28175
OY 180	LeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGly		199
Db 28174	TTGATTTTCATGATCACCAGGAAATTGTGTATACGGAAAGTACGCTACATTGGC		28115
OY 200	GlyArgAsnIleGlyAspGlyTrrPheLysValGlyGluAspThrValPheAlaAspLeu		219
	:::        :::		
Db 28114	GCGTTTATGTGTGCCGATCAATAT-----GCCGAGACACTAAAGCTTGGC-----		28067
OY 220	AspIleLeuAlaThrGlySerValValGlyGluValSerHisAspPheAspArgTyrTrp		239
	-----		
Db 28066	-----TATTGG 28061		
OY 240	AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly		259
Db 28060	CGGCGAT-----ACAAATTACCGATTCAAGGCCAGCCATCATTA 28019		
OY 260	LeuGln-----AlaLeuGlyTrrAsnAspGlyTrpIleSerArgHisAlaLeuLeuArg		276
	::::    :::		
Db 28018	CTGCAATGCGTTTGTATAGGATTTGGACGCTCTCCCGGAGAAAAATCGTGTGGC		27959
OY 277	TyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAsp		296
Db 27958	TAT-----CAATTGGATTAATTTCTTAAACTGAA--GCATTGGTGCA 27917		
OY 297	TrpGlnSerValGlnThrArgLeuLeuSerAspAspProAlaLysGlyLeuAspArgAsp		316
	-----		
Db 27916	GAGCAATATACATTCATTCAGATGATTCGCCAGTGCTC-----AACAGTACCGTGA 27863		
OY 317	ArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSer		336
	:::		
Db 27862	CAAAATTAATGGCTTATTAATTAATTAATTAATCTCTGCC-----AAGAAANA 27815		
OY 337	ValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLys		356
	:::		
Db 27814	GCTCGAATTCAAACCCCGATTTAGCTCTCGTAGTAGCTCTTGCGCTTTAAAGGTT		27755
OY 357	LeuValGlnAspGlyLysSerValThrValLeuThrAsnSerLeuGlnAlaThrAspVal		376
	:::		
Db 27754	GCTGGCGCTTCAGAGATGATGTTTAAATTAATGATTCCA-----GATAGCCCGACAT 27701		
OY 377	AlaAlaValHisSerGlyTrrValLysTyrArgLysProLeuLeuLysAlaGlyIleLys		396

Db	27700	CCGTTTATTTCGACGACACAGTATTACGCCCGCTTATTTCATGCAAAATAATTGCA	27641
Oy	397	LeuTYrGluLeuGlnProAsnHisAlaValProAlaThrLysAspSlyLeuThrgly	416
Db	27640	ATTTAATTTATAACGGTGGCTTC-----	27617
Oy	417	SerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIle	436
Db	27616	-----TTTATTCATCGCAAAACAAATGATCATGGATGAGTCAAGTCTGCACAGTT	27572
Oy	437	GlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValValIle	456
Db	27571	GGTTACGCCAAATCAACAGATATCGGAAGTTCACAAATTAAACTTTGAAAGCAAAATCCTGTGTTA	27512
Oy	457	GluSerProLysTlleAlaGluGlnMetGlu	466
Db	27511	TATGATAAAAAAATCATGTGATCAATTAGAA	27482
RESULT	27		
AA67781			
ID	AA67781	standard; DNA; 1389 BP.	
XX	AA67781;		
XX			
DT	11-APR-2001	(first entry)	
XX			
DE	Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:77.		
KW	Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;		
KW	membrane construction and membrane transport protein; petroleum spill;		
KW	hydrocarbon degradation; gram positive aerobic bacterium; marker;		
KW	identification; microorganisms; fine chemical production; transformation		
KW	genome mapping; genetic engineering; ds.		
XX			
OS	Corynebacterium glutamicum.		
XX			
PN	WO200100805-A2.		
PD	04-JAN-2001.		
XX			
PE	23-JUN-2000; 2000WO-IB00926.		
XX			
PR	25-JUN-1999;	990S-0141031.	
PR	08-JUN-1999;	99DE-1031454.	
PR	08-JUL-1999;	99DE-1031478.	
PR	08-JUL-1999;	99DE-1031563.	
PR	09-JUL-1999;	99DE-1032122.	
PR	09-JUL-1999;	99DE-1032124.	
PR	09-JUL-1999;	99DE-1032125.	
PR	09-JUL-1999;	99DE-1032128.	
PR	09-JUL-1999;	99DE-1032180.	
PR	09-JUL-1999;	99DE-1032182.	
PR	09-JUL-1999;	99DE-1032191.	
PR	09-JUL-1999;	99DE-1032209.	
PR	09-JUL-1999;	99DE-1032212.	
PR	09-JUL-1999;	99DE-1032222.	
PR	09-JUL-1999;	99DE-1032228.	
PR	09-JUL-1999;	99DE-1032229.	
PR	09-JUL-1999;	99DE-1032230.	
PR	14-JUL-1999;	99DE-1032927.	
PR	14-JUL-1999;	99DE-1033005.	
PR	14-JUL-1999;	99DE-1033006.	
PR	27-AUG-1999;	99DE-1040764.	
PR	27-AUG-1999;	99DE-1040765.	
PR	27-AUG-1999;	99DE-1040766.	
PR	27-AUG-1999;	99DE-1040830.	
PR	27-AUG-1999;	99DE-1040831.	
PR	27-AUG-1999;	99DE-1040832.	
PR	27-AUG-1999;	99DE-1040833.	
PR	31-AUG-1999;	99DE-1041378.	
PR	31-AUG-1999;	99DE-1041379.	

PR 31-AUG-1999: 99DE-1041395.  
PR 03-SEP-1999: 99DE-1042077.  
PR 03-SEP-1999: 99DE-1042078.  
PR 03-SEP-1999: 99DE-1042079.  
PR 03-SEP-1999: 99DE-1042080.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
XX P-PSDB: AAB76548.  
DR WPI: 2001-071486/08.  
XX  
XX  
PT Corynebacterium glutamicum nucleic acids encoding membrane construction  
PT and membrane transport proteins or their portions, useful for typing or  
PT identifying C. glutamicum or related bacteria, and as markers for  
PT transformation -  
XX  
XX  
PS Claim 3: Page 292-294; 111pp; English.  
XX  
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
CC construction and membrane transport (MCT) proteins given in AAB76510 to  
CC AAB76847. The MCT nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX  
XX  
SQ Sequence 1389 BP; 346 A; 518 C; 280 G; 245 T; 0 other;

## Alignment Scores:

Pred. No.: 4, 82e-15 Length: 1389  
Score: 243.50 Matches: 105  
Percent Similarity: 36.248 Conservative: 61  
Best Local Similarity: 22.938 Mismatches: 179  
Query Match: 8.988 Indels: 113  
DB: 22 Gaps: 17

US-10-066-551-4 (1-525) x AAF67781 (1-1389)

QY 70 ThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisGluAlaPheAlaIaArg 89  
DB 212 ACCGGCGGAACACGCGTTCTAC-----TCGACTACCGTGAATCCCTCAACGAGATG 265  
QY 90 AlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrPargAsn 109  
DB 266 ACCGGCGCATGACGACGACGAGAAATACATCTACGTGACGATCTACATCATGCGCTCG 325  
QY 110 AspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIaGluArgGlyValArg 129  
DB 326 GATTCCTACACCAACCATCTTTCGAGCAGCTGACGACGACCAACACGCGCGCTCAAA 385  
QY 130 ValArgLeuLeuLeuAsp----- 135  
DB 386 GTCCGCGCTCTTTTCGACGACGCTGGCAGCTGGAATACCCCGGCTACACACCGCTCAAA 445  
QY 136 ---AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsn 154  
DB 446 AAGAAGCTCAACCGCATGGGCTTCGCTGCTGACTCATG----- 484  
QY 155 IleGluValArgLeuPheAsnPropheValLeuArgIlyStrPargAlaLeuGlyTyrLeu 174  
DB 485 -----CTCCCGCTCAACCGCTGCGGA----- 505  
QY 175 ThrAspPheProArgLeuAsnArgArgMetHisAsnIlySerPheThrAlaAspAsnArg 194  
DB 506 CGCGCGCTTCGCGCGACCGCTGCGCAACGCGCAAAATGCTCATCATCGACGCGCAC 565  
QY 195 AlaThrIleLeuGlyGlyArgAsn---IleGlyAspGluTyr-----Phe 208  
DB 566 ACCGATTCATGGGCTCCCAAAATCTCATTCGCCGAGTTACTTACAAAAGAAAACATC 625  
QY 209 LysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValVal 228  
DB 626 AATCTGCGCGCGCA-----TGAAAGACCTCATGTGTCACTCAACCGCCCATCTGTC 679  
QY 229 GlyGluValSerHisAspPheAspArgTyrTyrPalaSerHisSerAlaHisAsnAlaThr 248  
DB 680 TCTCCATGGAATGATCTTCGCGCGAGCTGCTAGCTGCAATCCAAAC----- 727  
QY 249 ArgIleIleArgSerGlyAsnIleGlyIlyGlyLeuGlnAlaLeuGlyTyrAsnAspGlu 268  
DB 728 -----GAAAGCTCTGACATTCGCGCACAC 751  
QY 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288  
DB 752 GCAGAACCCGACGCGTACATCGGCACACACCAAAAGACTCCGCCACCAACCTCGTGCAG 811  
QY 289 LysIleGlnThrGlyArgIleAspTyrGlnSerValGlnThrArgLeuIleSerAspAsp 308  
DB 812 CTCATCCCTCCCGC----- 826  
QY 309 ProAlaIlyGlyLeuAspArgAspArgArgProProIleAlaIlyArgLeuGlnAsp 328  
DB 827 CCGTGTACACACACACCAACCTGCGC-----ATGTTCAACTCC 868  
QY 329 AlaLeuIlyGlnProGluIlySerValTyrLeuValSerProTyrPheValProThrIly 348  
DB 869 ATGCTTACACACCAACAGACATCTGTGAGCGCCCTCATTCATCCCGACGAA 928  
QY 349 SerGlyThrAspAlaLeuAlaIlySerValIleAlaValHisSerGlyTyrValIlyTyrArgIly 368  
DB 929 TCCCTCTGGAAACCGCTACCTGAGCTGACCGGAGTACCGTGAATCTATTCGTC 988  
QY 369 AsnSerLeuGlnIleThrAspValAlaAlaValHisSerGlyTyrValIlyTyrArgIly 388  
DB 989 TCT-----GAACAAACCGCAATTCGCTGAGTACGACGACCAATCTCTCTACAG 1042  
QY 389 ProLeuLeuIlyAlaGlyIleTyrLeuTyrGluLeuGlnProAsnHisAlaValProAla 408  
DB 1043 GCATCTCTGAAACCGCGGAGAAATCTACCAATTC-----CCCAACCCGACGTC 1093  
QY 409 ThrIlyAspIlyGlyLeuThrGlySerSerValThrSerLeuHisAlaIlyThrPheIle 428  
DB 1094 -----CTCCACACCAAGATACATGATC 1114  
QY 429 ValAsp-----GlyIlyArgIlePhe-----IleGlySerPheAsnLeu 441  
DB 1115 GCGGACCGCGACGACACACCGGCAACGAGCCCTGAGTCTCTCGATCTCTCAACCTC 1174  
QY 442 AspProArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProIlyIle 461  
DB 1175 GACATCCGAGCTTGTGGCTCACTACGAATCTCTGATGATGATCCCAAGGCAACCTC 1234  
QY 462 AlaGluGlnMetIlyArgThrIleuAlaAspThrSerProGluTyrAlaIlyArgValIthr 481  
DB 1235 ATCCACGACCTC---AACCGCTCACGACCGTTACCGACAGTAAAGTTCAAGCTCAC 1291  
QY 482 LeuAspArgHisAsnArgLeuGlnTyrPheAspProAlaThrArgIlyStrTyr 499  
DB 1292 TTGGATTAAGTGAACACGCGAGTTGG-----CGGCGCGCTAC 1330

## RESULT 28

AAH67979 standard; DNA; 1500 BP.

AAH67979;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 3014.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;



```

RESULT 29
AA168810
ID AA168810 standard; DNA; 1850 BP.
XX
AC AA168810;
XX
XX 25-JAN-2002 (first entry)
XX
DE C. glutamicum ATCC13032 cls DNA.
XX
XX Cardiolipin synthase; cps gene; coryneform bacterium; L-glutamate; ds;
XX amino acid production; human medicine; animal nutrition; food additive.
XX
XX Corynebacterium glutamicum.
XX
XX Key Location/Qualifiers
XX CDS 154..1656
XX FT /tag="a
XX FT /product="cls"
XX
XX DE10021826-A1.
XX
XX 08-NOV-2001.
XX
XX 04-MAY-2000; 2000DE-1021826.
XX
XX 04-MAY-2000; 2000DE-1021826.
XX
XX (DECS) DEGUSSA AG.
XX (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
XX Nampoochirt M, Moekkel B, Pfeifferle W, Eggeling L, Sahm H;
XX
XX WPI: 2002-027316/04.
XX
XX P-PSDB: AAG80237.
XX
XX New mutant coryneform bacterium, useful for production of amino acids,
XX especially glutamate, has increased activity of cardiolipin synthase
XX
XX Claim 9; Page 9-11; 20pp; German.
XX
XX This invention describes a novel genetically modified coryneform
XX bacterium (A) in which activity of the cardiolipin synthase (cls) gene,
XX encoding cardiolipin synthase, is increased. (A) are used for
XX fermentative production of amino acids, especially L-glutamate, useful in
XX human medicine, animal nutrition, pharmaceuticals and as food additives.
XX Nucleic acid, or its fragments, derived from the cls gene are used (1) as
XX primers for polymerase chain reaction of cls genes or (1i) as probes for
XX isolating cDNA and genes that are highly homologous with cls.
XX Overexpression of the cls gene results in increased production of amino
XX acids. This sequence encodes the Corynebacterium glutamicum ATCC13032
XX cps gene described in the method of the invention.
XX
XX Sequence 1850 BP; 447 A; 646 C; 396 G; 361 T; 0 other;
XX
Alignment Scores:
Pred. No.: 7,35e-15 Length: 1850
Score: 243.50 Matches: 105
Percent Similarity: 36.24% Conservative: 61
Best Local Similarity: 22.93% Mismatches: 179
Query Match: 8.98% Indels: 113
DB: 24 Gaps: 17
US-10-066-551-4 (1-525) x AA168810 (1-1850)
OY 70 ThrAsnGlyLeuSerAspIleTyrLeuLeuAsnProHisGluAlaPheAlaIaArg 89
DB 499 ACCGGGGAACACGCGTTTCTAC-----TCGACTACCGGTAAATCCCAACGATG 552
OY 90 AlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrArgAsn 109
DB 553 ACCGGCGCAATGACGACGAGCGAAGATACATCTACGTGAGATGTACATCATCATGCGCTGG 612

```

```

OY 110 AspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIaGluArgGlyValArg 129
DB 613 GATTCTACACCCACACCATTTCTTCGACGACCTGCAAGACGACCAACACCGCGCGTCAAA 672
OY 130 ValArgLeuLeuAsp----- 135
DB 673 GTCCGACTCTCTTTTCGACACACGTCGCGACGTGGAATATCCCGGTACACCGCTCAAA 732
OY 136 ---AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsn 154
DB 733 AAGAACTCAACCGCATGGCTTCGCTTCTGCTGATC----- 771
OY 155 IleGluValArgLeuPheAsnProPheValLeuArgLysTyrArgAlaLeuGlyTyrLeu 174
DB 772 -----CTCCCTCCCAACCTGGCGA----- 792
OY 175 ThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArg 194
DB 793 CGCGGCTTCGCGCGACCGACCTGCGCAACACCGCAAAATGCTCATCATCGACGCGCAC 852
OY 195 AlaThrIleLeuGlyLysArgAsn---IleGlyAspGluTyr-----Phe 208
DB 853 ACCGATTCATGAGGCTCCCAAAATCTCATCGCCCGAGTTACTACAAAAGAAACATC 912
OY 209 LysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValVal 228
DB 913 AAACCTGCGCGCGA-----TGGAAAGACCTCATGTGCACTACCGGCCCAACGTC 966
OY 229 GlyLysValSerHisAspPheAspArgTyrTyrPalaSerHisSerAlaHisAsnAlaThr 248
DB 967 TCCCTCATGAAATGATGCTTCGCGCGGACGTGTAGCTGATCAAC----- 1014
OY 249 ArgIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGlu 268
DB 1015 -----GAAGCCTCGACATCCGCGCAC 1038
OY 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288
DB 1039 GCAGAACCCACGCTACATCGGCAACACTCAAAAGACTCCGCGCAACCTCGTGCAG 1098
OY 289 LysIleGlnThrGlyArgIleAspTyrGlnSerValGlnThrArgLeuIleSerAspAsp 308
DB 1099 CTCATCCCTCGCGC----- 1113
OY 309 ProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyArgLeuGlnAsp 328
DB 1114 CCGGTTCACACACAGAACCCACCTGCGC-----ATGTTCAACTCC 1155
OY 329 AlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLys 348
DB 1156 ATCGTTCACACCGCCCAAAAGACATCATCTGTGCGACCTCTACTTCATCCCGCAGAA 1215
OY 349 SerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThr 368
DB 1216 TCCCTCTCGAAGCGCTACCTCAGCTCCTGACCGGAGTAGCTGACATATTCGTC 1275
OY 369 AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLys 388
DB 1276 TCT-----GAACAAACCGACCAATTCGCCATCGACCGCCCAATCTCTCACTACCGAG 1329
OY 389 ProLeuLeuLysAlaGlyLysLeuTyrGluLeuGlnProAsnHisAlaValProAla 408
DB 1330 GCACCTCTTGAAACCGCGCGAATCTACCAATTC---CCCAACCGCAGCTC----- 1380
OY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIle 428
DB 1381 -----CTCAGACCAAGTACATGATC 1401
OY 429 ValAsp-----GlyLysArgIlePhe-----IleGlySerPheAsnLeu 441
DB 1402 GCGAGCCCGGACGACACACCGGACGAGGAGGCTCGGATCTCGATCTCAACCTC 1461

```

QY 442 AspProArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerProLysIle 461  
DB 1462 GACATCCGCGTTGGCTGCAACTACGAAATCTCCGATGATGCCAAGGCAACCTC 1521  
QY 462 AlaGluGlnMetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThr 481  
DB 1522 ATCCACGAACTC---AAGCCCTCACCACCGTTACCGCAGCACTAGTTCACGCTCAC 1578  
QY 482 LeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLysThrTyr 499  
DB 1579 TTGGATTAAGTGGAGACACGCGCAGTTGG-----CGGCGCGCTAC 1617  
RESULT 30  
AAH68533  
ID AAH68533 standard; DNA; 349980 BP.  
AC AAH68533;  
XX  
XX 26-SEP-2001 (first entry)  
DE C glutamicum coding sequence fragment SEQ ID NO: 7068.  
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis; ds.  
XX Corynebacterium glutamicum.  
XX EPI108790-A2.  
XX 20-JUN-2001.  
PF 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX WPI: 2001-376931/40.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
XX  
XX  
PS Disclosure: SEQ ID NO: 7068; 246pp + Sequence Listing; English.  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX SO Sequence 349980 BP; 81250 A; 97718 C; 90621 G; 80391 T; 0 other;

## Alignment Scores:

Pred. No.: 1.63e-11 Length: 349980  
Score: 243.50 Matches: 105  
Percent Similarity: 36.24% Conservative: 61  
Best Local Similarity: 22.93% Mismatches: 179  
Query Match: 8.98% Indels: 113

DB: 22 Gaps: 17  
US-10-066-551-4 (1-525) x AAH68533 (1-349980)  
QY 70 ThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisGluAlaPheAlaIleArg 89  
DB 222463 ACCGGGGAACACACGCTTCTAC-----TCCGACTACCGTGAATCCCAACGAGATG 222516  
QY 90 AlaAlaLeuIleGluSerAlaGluHisSerLeuAspGlnGluTyrIleTyrPheAsn 109  
DB 222517 ACCGCCGACATCGACGAGGCGAGAAATACATCTACGTCGAGATCTACATGAGCTGG 222576  
QY 110 AspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleGluArgGlyValArg 129  
DB 222577 GATTCCTACACCCACCATCTCTTCGACGACCTCGAAGACGCCACAAACGCGGCGTCMAA 222636  
QY 130 ValArgLeuLeuLeuAsp----- 135  
DB 222637 GTCCGACTCTCTTTCGACGACGTCGCGCAGCTGGAGAAATACCCCGCTACACCGCTCAAA 222696  
QY 136 ---AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsn 154  
DB 222697 AAGAACTCAACCGCATGGGCTTCGCTGCTACCTCATG----- 222735  
QY 155 IleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeu 174  
DB 222736 -----CTCCCGCTCCACACCTGGCGA----- 222756  
QY 175 ThrAspPheProArgLeuAsnArgMetHisAsnLysSerPheThrAlaAspAsnArg 194  
DB 222757 CGCCGCTTCGCGACCGCCGACCGCCACCCGCAACCCGCAAAATGCTCATCATCGACGCCAC 222816  
QY 195 AlaThrIleLeuGlnLysArgAsp---IleGlyAspGluTyr-----Phe 208  
DB 222817 ACCGCAATTATGAGCTCCCAAAATCTATCGCCCGCAGTTCATACAAAGAAAACATC 222876  
QY 209 LysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValIle 228  
DB 222877 AACTCGGCGCGCA-----TCGAAAGACCTGATGTCGAACTCACCGGCCCATTCGTC 222930  
QY 229 GlyGluValSerHisAspPheAspArgTyrTrpAlaSerHisSerAlaHisAlaThr 248  
DB 222931 TCCTCCATGGAATGATCTCCGCGCGACCTGATCGAATCCAAAC----- 222978  
QY 249 ArgIleIleArgSerGlyLysIleGlyLysGluAlaLeuGlyTyrAsnAspGlu 268  
DB 222979 -----GAAGCCTCGACATCCGCGACAC 223002  
QY 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288  
DB 223003 GCAGAGCCCGACGGCTACATCGGCACACTCAAAAGACCTCCGCGCACCAACTCGTCGAG 223062  
QY 289 LysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAspAsp 308  
DB 223063 CTCATCCCTCCGCGC----- 223077  
QY 309 ProAlaLysGlyLeuAspArgAspArgLysProProIleAlaLysArgLeuGlnAsp 328  
DB 223078 CTGGTTACACACACAGACCAACCTGCGC-----ATGTTCAACTCC 223119  
QY 329 AlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLys 348  
DB 223120 ATCGTTACACCGCCCAAGACAGACATCTGTGAGCCCGCTACTTCACCGCGAGAA 223179  
QY 349 SerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThr 368  
DB 223180 TCCCTTCCTGGAAGCCCTCACTCAGCCCTCTACCGCGGAGTAACCTCGACATTCCTC 223239  
QY 369 AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLys 388  
DB 223240 TCT-----GAACACGCGACCAATTCGCGATCGACCGCCCAATCTCTCTACTACCG 223293  
QY 389 ProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAla 408

```
Db 223294 GCACCTCCTGAAGCCGCGCTGAATAATCTACCAATTCC--CCCAAAACCCGACGCTC----- 223344
QY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaThrPheIle 428
Db 223345 -----CTCCACACCAAGTACATGATC 223365
QY 429 ValAsp-----GlyLysArgIlePhe-----IleGlySerPheAsnLeu 441
Db 223366 GCCACCGCCGACGACACACCGGCAAGACCGCTCGAGTCTCGATCCTCCACACCTC 223425
QY 442 AspProArgSerAlaArgLeuAsnThrGluMetGlyValIleLeuSerProLysIle 461
Db 223426 GACATCCGCGCTTGGCTCCCAACTACGAATCTCCGTATGTCGCCAAAGGCAACCTC 223485
QY 462 AlaGluMetGluArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThr 481
Db 223486 ATCCACGAACTC---AACCCTCCTACCCGACCGCTTACCGACAGTAACTTCAAGCTCACC 223542
QY 482 LeuAspArgHisAsnArgLeuGlnThrPheHisAspProAlaThrArgLysThrTyr 499
Db 223543 TTGGATAACTGGACACGCGCAGTTGG-----CGCGCGCGCTAC 223581
```

## RESULT 31

AAS59588 standard; DNA; 18096 BP.

AAS59588;

13-FEB-2002 (first entry)

Propionibacterium acnes immunogenic protein encoding DNA #83.

SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
dermatological; osteopathic; neuroprotectant; ds.

Propionibacterium acnes.

WO200181581-A2.

01-NOV-2001.

20-APR-2001; 2001WO-US12865.

21-APR-2000; 2000US-199047P.

02-JUN-2000; 2000US-208841P.

07-JUL-2000; 2000US-216747P.

(CORI-) CORIXA CORP.

Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

L'maisonneuve J, Zhang Y, Jen S, Carter D;

WPI; 2001-616774/71.

Propionibacterium acnes polypeptides and nucleic acids useful for  
treating acne vulgaris -

Claim 1; SEQ ID No 83; 1069pp; English.

Sequences AAS59506-AAS59804 represent DNA molecules encoding  
acnes immunogenic polypeptides. The proteins and their  
associated DNA sequences are used in the treatment, prevention and  
diagnosis of medical conditions caused by P. acnes. The disorders include  
SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
in infections of bone, joints and the central nervous system, however it  
is particularly involved in the inflammatory lesions associated with acne  
vulgaris. A method for detecting the presence or absence of P. acnes in a  
patient comprises contacting a sample with a binding agent that binds to

CC the proteins of the invention and determining the amount of bound protein  
CC in the sample. The polypeptides may be used as antigens in the production  
CC of antibodies specific for P. acnes proteins. These antibodies can be  
CC used to downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
CC polypeptides shown in AAU57920-AAU58104 and AAU67598.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 18096 BP; 3826 A; 5738 C; 5088 G; 3444 T; 0 other;

## Alignment Scores:

Pred. No.:	3e-13	Length:	18096
Score:	242.00	Matches:	117.
Percent Similarity:	36.388	Conservative:	58
Best Local Similarity:	24.328	Mismatches:	152
Query Match:	8.928	Indels:	155
DB:	23	Gaps:	20

US-10-066-551-4 (1-525) x AAS59588 (1-18096)

```
QY 29 LeuLeuLeuCysSerCysSerSerTrp-----LeuProProLeuGluLysArg 44
Db 8079 CTTTCGCCCGCCGACCTGCTGGTGGCGGAGTACTGCTCCGCCGATGGCCAGCA 8138
QY 45 -----ThrGluSerArgHisPheAsnThrSerLys 54
Db 8139 TGGTCGGACTACTACCTTCAACGGCATTAACGACGAGCGCGTCCCATCCGTTCC 8198
QY 55 ProValArgLeuAspAsnIleLeuGlnIleArgHisThrPro----- 68
Db 8199 CCACTCCGC-----CCACGAGCCCTTCAAGCTGTGTGTCATC 8237
QY 69 ---HisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisGluLysPheAla 87
Db 8238 AGCTCGAGATCTTCCGCTTGGACGTGAT-CTT-----TACGCG 8275
QY 88 AlaArgAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTrp 107
Db 8276 GCGATGATGCGCGACATGACGTACGTACACACCGCTTACTTGAAGCTTCATCTGG 8335
QY 108 ArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleArgGly 127
Db 8336 AAAAAGCAGCAGAGTGGCGCTATTTCGCGAGCGACTGTGAGAGCCGCCGACGCGGA 8395
QY 128 ValArgVal-----ArgLeuLeuLeuAspAsnAsnThr 139
Db 8396 GTCAAGCTTATGCGATGTGGGACACCTTCCGCCACCTGTTGTGGACCCACGATTTCTT 8455
QY 140 ArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeu 159
Db 8456 CGGCACCTCGAG-----GGGCTGCGAGTGGCAGGC 8485
QY 160 PheAsnProPheValLeuArgLysTrp-----ArgAlaLeuGlyTyrLeuThr 175
Db 8486 ---CAACTCTGTACTCTCCATCATGATACGACGCTACGAATCTGGGG----- 8533
QY 176 AspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195
Db 8534 -----CGAGACACCGCCCAACCTCTCATCTCGTCGATCGTCAAAAGTC 8572
QY 196 ThrIleLeuGlyArgAsnIleGlyAspArgLysTyrPheLysValGlyLysPheVal 215
Db 8573 GCGTACATCGGTGGGTACACATCGCTCTTACGCGGACCGCTGGCGTACAC--- 8629
QY 216 PheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyLysSerHisAspPhe 235
Db 8630 -----CATGCTGTATTACGGGCGCTGCGGTGGGAGTGCAGAACGCTTCTT 8677
QY 236 AspArgTyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsn 255
```



```

Db 8678 GTTGACATGTGG----- 8689
QY 256 ILeIlySGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeu 275
Db 8690 -----AATCAACGCCCGCAAGGAGGCGCTATC 8716
QY 276 ---ArgTyrArgGluThrValGlnGlnSerProLeuTyrGlnLysIleGlnThrGlyArg 294
Db 8717 CCCCCGGGCAACGAGCCCGCTACCCATCC-----GCCGTC 8755
QY 295 ILeAspTyrGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAsp 314
Db 8756 CGCATATTGG-----GATACCGGCTTTCGCTGTACAC 8785
QY 315 ArgAspArgArgLysPro-----ProIleAlaGlyLeuGlnAspAlaLeuLys 331
Db 8786 CGCAATTCCCCCGGATGGCGGTCTATCCATCCGGAACATGTACTGGAGGCGCAT 8845
QY 332 GlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThr 351
Db 8846 CGCCCGAGTACGCGATCTGTGATACCCAGGCTACTCTATCCCCGATGACGAGCTTGT 8905
QY 352 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeu 371
Db 8906 GACACCTCCACCAAGCTGATCTCGGGGGTGTAGATC----- 8950
QY 372 GlnAlaThrAspValAlaAlaValHisSerGlyTyrVal----- 384
Db 8951 -----GTCATCCGCGTGAATCGAACCATCATCGCCGACGCTGACGCCG 8998
QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHis 404
Db 8999 GGCATTACGACCGCTCTCTCTACTAGGCTGGCGCTTCTTACCAG----- 9049
QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424
Db 9050 -----GGAGCTATGGTC-----CACGCG 9067
QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444
Db 9068 AAGACCTGCACTATTATGAGGATTTGGTCGACGATTTGTACAGCGAACCTTGACCGGCTA 9127
QY 445 SerIlaArgLeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluGln 464
Db 9128 TCCCTCAAGTACTATGAGTAAGCTGCTTCACTGACTGACCGCGCCCAACGA 9187
QY 465 Met 465
Db 9188 ATG 9190

RESULT 32
AAH52803
ID AAH52803 standard; DNA; 1325 BP.
XX
AC AAH52803:
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:999.
XX
KM Staphylococcus epidermidis SRI strain; infection; diagnosis:
XX
OS Staphylococcus epidermidis.
XX
PN MO200134809-A2.
XX
PD 17-MAY-2001.
XX
PE 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 9905-0164258.

```

```

XX
PA (GLAX ) GLAXO GROUP LTD.
XX
PI Kimmberly WJ;
XX
DR WPI: 2001-316495/33.
XX
DR P-PSDB: AAG81953.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 291-292; 2188pp; English.
XX
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4455 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 1325 BP; 453 A; 182 C; 239 G; 451 T; 0 other:
XX
Alignment Scores:
Pred. No.: 2,1e-14 Length: 1325
Score: 237.00 Matches: 89
Percent Similarity: 37.60% Conservative: 58
Best Local Similarity: 22.76% Mismatches: 156
Query Match: 8,74% Indels: 88
DB: Gaps: 9
US-10-066-551-4 (1-525) x AAH52803 (1-1325)
QY 76 ILeTyrLeuLeuAsnAspProHisGlnAlaPheAlaAlaLeuIleGluSer 95
Db 385 ATCGATTATTATTACAGATGGTCATTAAGCTTATGAAAAAGTACTTGAGATATTACAT 444
QY 96 AlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrPArgAsnAspIleSerGlyArgLeu 115
Db 445 GCTCAAGACTATATACATCTAGACTATATACCTTTGAACTTGATGATTAAGTAAAGA 504
QY 116 LeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAsp 135
Db 505 ATCTTAGATGCACTTGAACAACTAACTTAAGAGCTTAAAGAGTTAACTTTGTATGAC 564
QY 136 AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155
Db 565 GATGTGGTCTTAA----- 579
QY 156 GluValArgLeuPheAsnProPheValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThr 175
Db 580 AAGGTTAATTATCAAA-----TTTAAACATTTCAGAGCATTTAGTGGAACATT 630
QY 176 Asp-----PhePro-----ArgLeuAsnArgArgMetHis 185
Db 631 GAAGCATTTTCCCTCGAAAGTACTTTAATCAATTCAGATGAATAAATCAAGAAATCAT 690
QY 186 AsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAsp 205
Db 691 AGAAGATTATTCATTATGATGAGCAAAATGGTTAGTGGCGGTTTAAATGTGCGCAT 750

```

```

QY 206 GltuTyRheleValGlyLysPThrValPheAlaAspLeuAspIleuAlaThrGly 225
DB 751 GATTATTAGTAGTAAAGTAAAGTTACTGAGAGATACATACAGCTGTTCAAGT 810
QY 226 SerValValGlyLysValSerHisAspPargTyrTrpAlaSerHisSerAlaHis 245
DB 811 GAATGCATCATGCACATCAATTAAGATTATTTTACAGGAAATTCACATGCG- 864
QY 246 AsnAlaThrArgIleLeuArgSerGlyAsnIleGlySerGlyLeuAlaLeuGlyTyr 265
DB 864 ----- 864
QY 266 AsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerPro 285
DB 865 -----CATGTCACAAATTTAATTTGATCA 891
QY 286 LeuTyrGlnLysIleGlnThrArgIleAspTrpGlnSerValGlnThrArgLeuIle 305
DB 892 AAATATTTCCCTAAAAAATGCG- -GACAAAGGAACGGCGCTAATTCAAATCGCT 945
QY 306 SerAspAspProAlaLysGlyLeuAspArgAspArgTyrProIleAlaGlyArg 325
DB 946 TCAAGGAGACTGCTGATTGATTACATCA- - - - -ATAGCATATGCT 987
QY 326 LeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheVal 345
DB 988 TATACAAAAATGATATGACCGCTAAAAAACTATATCTATCTACAAACCCCTTACTTAT 1047
QY 346 ProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThr 365
DB 1048 CCGACCAATCATACATTAATGATTAATAAATGGCTCTAATACGGCGCTTGAAGTAAC 1107
QY 366 ValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLys 385
DB 1108 CTTATGATACCGTGT- - - - -AAACCTGATCATCATCTGTTTATGGGTACATTTCA 1161
QY 386 TyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluGlnProAsnHisAla 405
DB 1162 AATGACAGCTGATTATTTGATGACGGAGTTAATTTACACTTATCAAAATGATTT- 1218
QY 406 ValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLys 425
DB 1219 -----ATTCATTTCTAAA 1230
QY 426 ThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSer 445
DB 1231 ATATTATATGATTGATGATGAATTTCTTCATTTGTTGTCACAAACATGACTTTAGAGC 1290
QY 446 AlaArgLeuAsnThrGluMetGlyValIle 456
DB 1291 TTGAACTGAATTTTCGAAGTGAATGATGATTTATA 1323
RESULT 33
AAH54735/c
ID AAH54735 standard; DNA; 2978 BP.
AC AAH54735;
DT 03-SEP-2001 (first entry)
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:4099.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.

```

```

XX 09-NOV-1999; 99US-0164258.
PR
PA (GLAXO ) GLAXO GROUP LTD.
XX
XX kimmerly WJ;
XX
XX WPI; 2001-316495/33.
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 8; Page 1789-1790; 2188pp; English.
PS
XX AAH53304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AG81454 to AG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH5090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH5091 to
XX AAH5098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
SQ Sequence 2978 BP; 987 A; 474 C; 464 G; 1053 T; 0 other;
XX
Alignment Scores:
Pred. No.: 6,91e-14 Length: 2978
Score: 237.00 Matches: 89
Percent Similarity: 37.60% Conservative: 58
Best Local Similarity: 22.76% Mismatches: 156
Query Match: 8,74% Indels: 88
DB: Gaps: 9
US-10-066-551-4 (1-525) x AAH54735 (1-2978)
QY 76 IletYrLeuLeuAsnSproHisGluAlaPheAlaAlaArgAlaLeuIleGluSer 95
DB 954 ATCGATTATTATTACAGATGGCTATTAAGCTTTATGAAAAAGTACTTGAGATATTATCAAT 895
QY 96 AlaGluHisSerLeuAspLeuGlnTyrTyrIleTrpArgAsnAspIleSerGlyArgLeu 115
DB 894 GCTCAAGACTATATACATCATCTAGAGTACTTACTTGAACCTTGATGATTAGTAAGA 835
QY 116 LeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAsp 135
DB 834 ATCTTAGATGCACCTTGAACTTAAGAAAGAGTTTAAAGCTTAAACTTTTGTATGAC 775
QY 136 AspAsnAspThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155
DB 774 GATGTTGGTCTTAA- - - - - 760
QY 156 GluValArgLeuPheAsnPropheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThr 175
DB 759 AAGGTAGATTATCAAAA- - - - -TTTAAACATTTTCAGACATTTAGTGAGAGATT 709
QY 176 Asp- - - - -PhePro- - - - -ArgLeuAsnArgArgMetHis 185
DB 708 GAAGCATTTTCCCTTGAAGAGTACCTTTAATCAATTTTCAGATGAATATCAAAATCAT 649
QY 186 AsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyLysArgHisIleGlyAsp 205
DB 111 ----- 111

```

```

Db 648 AGAAGATATATCATATAGATGACAAATGCTTACGCTGGCTTTATGTCGGCAT 589
Qy 206 GltutyrthelysValIgluIaspThrValPhehlaaspleuasprIleuAlaIthGly 225
Db 588 GATTATTTAGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 529
Qy 226 SerValValIgluIaspSerHisPheasparGlyTyrPalaSerHisSerAlaHis 245
Db 528 GAATGATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 475
Qy 246 AsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyr 265
Db 475 ----- 475
Qy 266 AsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerPro 285
Db 474 -----CATGCTCCACAATTTAAATTTGATCA 448
Qy 286 LeuTyrGlnLysIleGlnThrGlyArgIleAspThrGlnSerValGlnThrArgLeuIle 305
Db 447 AAATATTTCCCTAAATAAATGCG-----GACAAGGAAACGGCGCATTCATCAATCGCT 394
Qy 306 SerAspAspProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyArg 325
Db 393 TCTAGTGGACCTGATTTGATTTACATCA-----ATGGAATATGCT 352
Qy 326 LeuIlnsAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheVal 345
Db 351 TATACAAAATGATTAATGAGCGCTAAAGTCTATCTATCTACAAAGCCCTTACTTATTT 292
Qy 346 ProThrLysSerGlyThrAspAlaLysLeuValGlnAspGlyIleAspValThr 365
Db 291 CCACAGCAATCATCATTAATGATTAATAAATGCGCTGATTAAGCGCGGTGAAGTAAAC 232
Qy 366 ValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLys 385
Db 231 CTATGATACCGTGT-----AACCTGATCATCATTCCTTATTTGGGCTACATTTTCA 178
Qy 386 TyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAla 405
Db 177 AATCAGACTGATTTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
Qy 406 ValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLys 425
Db 120 -----ATTCATTTCTTAA 109
Qy 426 ThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspArgSer 445
Db 108 ATATTTAATGATTTGATGATGATGATTTCTTCAATGCTAGTGCACAACATGACATTTAGAAC 49
Qy 446 AlaArgLeuAsnThrGluMetGlyValValIle 456
Db 48 TTTCAGACTGAATTTTCAAGTGAATGATCATTTATA 16

```

RESULT 34  
AAH54681/c  
ID AAH54681 standard; DNA; 3745 BP.

AC AAH54681;

DT 03-SEP-2001 (first entry)

DE 5. epidermidis genomic polynucleotide sequence; SEQ ID NO:4045.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;

OS Staphylococcus epidermidis.

XX MO200134809-A2.

XX 17-MAY-2001.

XX

PF 09-NOV-2000; 2000MO-US30782.  
XX  
PR 09-NOV-1999; 99US-0164258.  
XX  
PA (GLAX) GLAXO GROUP LTD.  
XX  
PI Kimmberly MJ.  
XX  
DR WPI; 2001-316495/33.  
XX  
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
XX  
PS Claim 8; Page 1727-1728; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
CC (I) and (II) can have antibacterial activity and therefore can be used  
CC in vaccination. The nucleic acids (I) may be used to produce the  
CC S. epidermidis polypeptides (II) via the production of vectors  
CC containing them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH5091 to  
CC AAH5098 represent oligonucleotide sequences and primers which are used  
CC in the exemplification of the present invention.  
CC N.B. The present invention specifically claims all the polynucleotide  
CC sequences given in the sequence listing of the present specification,  
CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
CC though sequences are present in the disclosure for SEQ ID NO:4455 to 4472,  
CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 3745 BP; 1352 A; 610 C; 568 G; 1215 T; 0 other;

Alignment Scores:  
Pred. No.: 9,68e-14 Length: 3745  
Score: 237.00 Matches: 89  
Percent Similarity: 37.60% Conserved: 58  
Best Local Similarity: 22.76% Mismatches: 156  
Query Match: 8.74% Indels: 88  
DB: Gaps: 9

US-10-066-551-4 (1-525) x AAH54681 (1-3745)

```

Qy 76 IleTyrLeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSer 95
Db 941 ATCGATTATTTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 882
Qy 96 AlaGlnHisSerLeuAspLeuGlnTyrIleTyrPArgAsnAspIleSerGlyArgLeu 115
Db 881 GCTCAGACTATATACATCATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 822
Qy 116 LeuPheAsnLeuValTyrLeuAlaAlaGluIArgGlyValArgValArgLeuLeuAsp 135
Db 821 ATCTTAGATGACCTTGGAACACTTAAGAGAGTTTGAAGAGTTTGAAGAGTTTGAAGAG 762
Qy 136 AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155
Db 761 GATGTTGGTCTTAAA-----TTTAAACATTTTCAAGCATGATGATGATGATGATGATGAT 747
Qy 156 GluValArgLeuPheAsnProPheValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThr 175
Db 746 AAGGTTAGATTATCAAAA-----TTTAAACATTTTCAAGCATGATGATGATGATGATGATGAT 696
Qy 176 Asp-----PhePro-----ArgLeuAsnArgArgMetHis 185
Db 695 GAAGCATTTTCCCTTGCAGAAAGTACCTTTAATCATTTTCAGATGATGATGATGATGATGATGAT 636
Qy 186 AsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAsp 205

```

```

Db 635 AGAAGATTATCATTTATGATGCAAAATTTGGTTACGTTGGCGTTTATGTCGGCGCAT 576
QY 206 GUTYrPhelYsValGlyGlnSphrValPheAlaSpLeuAspLleuAlaThrGly 225
Db 575 GATTATTGGATTAGTAGTAAAGTTAGTTACTGAGAGATACACATACAGCTGTCAAGGT 516
QY 226 SerValValGlyGlnValSerHisAspPheAspArgTyrTrpAlaSerHisSerAlaHis 245
Db 515 GATGCGATGATGACATCAATTAAGATTATTTAGTACTGAAATTCACAGTCG----- 462
QY 246 AsnAlaThrArgLleLeuArgSerGlyAsnLleGlyLysGlyLeuGlnAlaLeuGlyTyr 265
Db 462 ----- 462
QY 266 AsnAspGluThrSerArgHisAlaLeuAlaLeuAlaArgTyrArgGluThrValGlnGlnSerPro 285
Db 461 -----CATCGTCCACCAATTTAAATTTGATCA 435
QY 286 LeuTyrGlnLysLleGlnThrGlyArgLysAspTrpGlnSerValGlnThrArgLeuLle 305
Db 434 AATATTTCCTCAAAAAAATGCG-----GACCAAGCAACGCGCTATTCAATCGCT 381
QY 306 SerAspAspProAlaLysGlyLeuAspArgAspArgLysProProlleAlaGlyArg 325
Db 380 TCTAGTGCACCTGCATTGATTATTCATCA-----ATGAAATATGCT 339
QY 326 LeuGlnAspAlaLeuLysGlnProGlnLysSerValTyrLeuValSerProTyrPheVal 345
Db 338 TATCAAAAAATGATATGACCGCTAAAGTCTATCTATCTACAAAGCCCTTACTTATT 279
QY 346 ProThrLysSerGlyThrAspAlaValLysLeuValGlnAspGlyLysValThr 365
Db 278 CCAGACCATCATACATTAATGCAATTAATAATGCGCTGAATGCGGCTTGAAGTAAC 219
QY 366 ValLeuThrAsnSerLeuGlnAlaThrAspValAlaValHisSerGlyTyrValLys 385
Db 218 CTATATATACCGTGT-----AAACCTGATCATCATTCGTTATTTGGGCTACATTTTCA 165
QY 386 TyrArgLysProLeuLysAlaGlyLysLysLeuTyrGlnLeuGlnProAsnHisAla 405
Db 164 AATGACGCTGATTATTGATGACGAGGATTAATATTACCTTATCAAAATGATTT--- 108
QY 406 ValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLys 425
Db 107 -----ATTCATTTCTAAA 96
QY 426 ThrPheLleValAspGlyLysArgLlePheLleGlySerPheAsnLeuAspProArgSer 445
Db 95 ATATTATGATGATTGATGATAAATTTCTTCAATTGGTAGTGCAACAATGACCTTTAGAACG 36
QY 446 AlaArgLeuAsnThrGlnMetGlyValValLle 456
Db 35 TTGGAAGTGAATTTGACAGATGATGATTAATA 3

```

```

PN WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US16450.
XX 19-MAY-2000; 2000US-205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
PI Birse CE, Rosen CA;
XX WPI: 2002-122018/16.
DR P-PSDB; ABB90259.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
PS Claim 4: SEQ ID NO 1230; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABL90668) and proteins
CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1877 BP; 426 A; 523 C; 485 G; 436 T; 7 other;

```

```

Alignment Scores: 2.51e-12 Length: 1877
Pred. No.: 219.00 Matches: 86
Score: 35.20% Conservative: 52
Percent Similarity: 21.94% Mismatches: 164
Best Local Similarity: 8.07% Indels: 90
Query Match: 24 Gaps: 9

```

US-10-066-551-4 (1-525) x ABL90668 (1-1877)

```

QY 76 LLeTyrLeuLysAsnAspProHisGluAlaPheAlaAlaArgAlaLeuLleGlnSer 95
Db 995 ATCCAGTTCGCGAANAAGCGGACGACATATATATCCCGGCTTTAAAGCGATGGCGAG 936
QY 96 AlaGlnHisSerLeuAspLeuGlnTyrTyrLleTrpArgAsnAspLysSerGlyArgLeu 115
Db 935 GCACAGAACGCAATCATTTCTGTAAGCTTTATCTGTTGAGATGACCGCGCAACAA 876
QY 116 LeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAsp 135
Db 875 CTGCAATGCGGCACTACTGCGACGACGCAACCGCGGGTTAAACGCAAGCTTGTGAT 816
QY 136 AspAsnAsnThrArgLysLeuAspPheLeuLeuAlaLeuAspSerHisProAsnLle 155
Db 815 GCGTACGGTTCGCGGATTCACGCGATGAGTTGTCAATGAAGTGAACGACGCTGCGGTA 756
QY 156 GluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThr 175
Db 755 GTGTTCGCTACTACATGCC-----CGCCCTCGCCTTTTGGTATGCGCACG 708

```

```

RESULT 35
ABL90668/C
ID ABL90668 standard; cDNA; 1877 BP.
XX
XX ABL90668;
XX
XX 24-MAY-2002 (first entry)
XX
XX Human polynucleotide SEQ ID NO 1230.
XX
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX antiatherogenic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX antitumor; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.
XX
XX Homo sapiens.
XX

```

```

QY 176 AspPheProArgLeuAsnArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195
Db 707 AAT-----GTGTTCCCGGATGATCGCAAAATTTGGTGATCGACGCCGATATA 657
QY 196 ThrIleuGlyGlyArgAsnIleGlyAspGlyuTrpPheLysValGlyGluAspThrVal 215
Db 656 GCCTTTATTTGGCGGCTGAATTACTCCCGGATATGTCACACTACGTCAGAGAGCT 597
QY 216 PheAlaAspLeuAsnIleLeuAlaThrGlySerValAlaGlyGluValSerHisAspPhe 235
Db 596 AAGAGGATTAACGGGATACCGCTTGAAGGCCGATTTGTCCAGATATTTCTC---CAGTTT 540
QY 236 Asp-----ArgTyrTrpAlaSerHisSerAla 244
Db 539 GAGCTGGAACAACTGCTGGACAGAGCGGCGACAGCGGTGGTGGCGACGTCAT----- 486
QY 245 HisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeuAlaLeuGly 264
Db 485 CACAAACCGGAGAGAACCCCGACCGGGA----- 456
QY 265 TyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSer 284
Db 455 -----GAGCGCCAGGTATTTGCTGCTGCTGCGC----- 429
QY 285 ProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeu 304
Db 429 ----- 429
QY 305 IleSerAspAspProAlaLysGlyLeuAspArgAspArgArgLysProPheIleAlaGly 324
Db 428 -----GATTAACGAGAACACATCCGATGATATTAAGACGC 396
QY 325 ArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPhe 344
Db 395 CATATTGGAAGATGCTCACCACGCGCGGCGGAGAGTATATCCCAACGCTACTTTC 336
QY 345 ValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364
Db 335 TTCCCGCGCATTCGATTTTATACGCTTGGCTAAAGCGGCGCGGCGGATC 276
QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrVal 384
Db 275 AAACGTGATCATTCAGGC-----GAACCGGATATGCCGATTTGACAGAGTGGCGCGC 222
QY 385 LysTyrArgLysProLeuLeuLysAlaGlyTyrLeuLysTyrGluLeuGlnProAsnHis 404
Db 221 TTGCTGATATACATCTCTGGTTAAAGCGCGCTTACAGTTTTCGATACCGCGCCG--- 165
QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424
Db 164 -----CCGCTCCACGCG 153
QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444
Db 152 AAAGTGGCATTTGATGAGCATCTAGTGGCGGACACATAGGCTCAGTAATCTGATCCGCTC 93
QY 445 SerAlaArgLeuAsnThrGluMetGlyValValIle 456
Db 92 AGTTTGTCATGATCTCGAAGCAATGTCATCATC 57

```

```

RESULT 36
ABA92787/c
ID ABA92787 standard; DNA: 640681 BP.
XX
AC ABA92787;
XX
DT 27-MAR-2002 (first entry)
XX
DE Buchnera sp. genomic DNA SEQ ID NO:1.
XX
KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
circular; ds.
XX

```

```

OS Buchnera sp.
XX
PN JP2001292771-A.
XX
PD 23-OCT-2001.
XX
PF 07-APR-2000; 2000JP-0107160.
XX
PR 07-APR-2000; 2000JP-0107160.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
WP: 2002-126043/17.
XX
PT A genomic DNA of cockroach-symbiotic bacterium -
PS Claim 1; Page 16-230; 237pp; Japanese.
XX
CC The present invention describes a gene (I) derived from Buchnera sp.
CC containing the DNA (a) or (b), (a) has a fully defined base pair
CC sequence selected from a table of sequences found in the Buchnera sp.
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (I); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC or (d), (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and
CC the expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents
CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
XX
SQ Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1,8e-07 Length: 640681
Score: 208.00 Matches: 105
Percent Similarity: 37.88% Conservative: 81
Best Local Similarity: 21.38% Mismatches: 163
Query Match: 7.67% Indels: 142
DB: Gaps: 21
XX
US-10-066-551-4 (1-525) x ABA92787 (1-640681)
QY 49 HisPheAsnThrSerLysProValArgLeu-----AspAsnIleLeuGlnIle 64
Db 299359 CACGGAATTAAAGCAGAAATAATTAATACTAATACCAAGAAATAATATGCAATA 299300
QY 65 ArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisIle 84
Db 299299 TTA-----ATACGTAGTATTATTTTA----- 299279
QY 85 AlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyr 104
Db 299278 -----GCACGTAAATAATTTGAATGTTTTT 299252
QY 105 TyrIleTrpArg-----AsnAspIleSerGlyArgLeuLeuPheAsnLeu 119
Db 299251 TATATTGGAACACAGCGTGGATGCGATGATGTGGCAATAGCTGATTTGAT----- 299198
QY 120 ValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAspAspAsnThr 139
Db 299197 -----TCTGCAAAACGTTGATATCATTTGATTAATGCTTGATTCAGCAGAAAGT 299147
QY 140 ArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeu 159
Db 299146 -----ATTGAATTTTTCAGACCCCTGGGTGGAATATATGAGA 299108
QY 160 PheAsnProPheValLeuValArgLysTrpArgAlaLeuGlyTyrLeuThrAspPhe---Pro 178

```





## Alignment Scores:

Pred. No.:	1,47e-10	Length:	933
Score:	197.50	Matches:	86
Percent Similarity:	34.99%	Conservative:	48
Best Local Similarity:	22.45%	Mismatches:	148
Query Match:	7,288	Indels:	101
	24	Gaps:	11

US-10-066-551-4 (1-525) x ABK73802 (1-933)

```

QY 124 ALGLIURGGLYVALARGVALARGLEULEULEASPASPSANSTHARGGLYLEUASP 143
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12 GCGAAATCCGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATG 71
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 144 ASPLEULEULEALEUASPSEPHISPROASNILEGLUVALARGLEUPHESPHE 163
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 72 AAAAAACGCTCCGCCGCGAAACAAAGCGGTGCGACGTTTTCGCCAAC----- 125
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 164 VALLEUARGYSTPRARGALALEUGLYTYRLEUTHASPHEROARGLEUASARG 183
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 126 -----AAGCGCGCTCCGCCATTTCTTTTACCGCGTGAATGCGAG 167
    |||::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 METHISANLYSSERPHETHRALASPASNARGALATHRILEUGLYARGASNILE 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 168 AATCATGAAATAATCCGCCGATGACGGAAGATCGGCTATGCGCGATTTAAACATC 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 204 GLYASPGIUTYRPE--LYSVALGLYGLUASPRHVALPHEALAASPLEUASPILEU 222
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 228 GCGAAGAAATACCTTGTAATAAAAGCGAATTCGACCGGAAGATTTACATTTCGA 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 223 ALATHRGILYSERVALVALGLYGLUVALSERHIS-----ASPHESPARGTYR 238
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 288 ATGACGCGGAAGAGCGCTGCGACACTGACACATTTTATATGCGACTTCAAA----- 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 239 TRPALASERHISSERALAHISNALATHARGILEILEARYSERGLYASNILEGLY 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 341 ----- 341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 GLYLEUGINALALEUGLYTYRASNAPGLUTHRSERARGHISALALEULEUARGTYR 278
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 342 -----AGA 344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 GLUTHRALGLUGINSERPRO-----LEUTYRGLINLSILEGINTHRGYARGILE 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 345 GAAGCGCGCGGCAAGGCGCGCACAGCGATTTCCGCCGCGACGAAAGGCGTCTC 404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 296 ASPTRGLINSERVALGINTHARGLEUILESERFASPRPROALALYSGLYLEUASPAR 315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 405 ACCCATACAAACCATGCGACAAAGGATTTCT----- 437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 316 ASPARGARGLYSPROPROILEAGLYARGLEUGINALALEUYSGLINPROGLULYS 335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 438 -----CTAAGAGAAATAATACATTTCATTATGCAACGCGAAAGAA 479
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 SERVALTYRLEUVALSERPROTYRPHREVALPROTHRYSERGLYTHRASPALALEUAL 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 480 AGAATGCTGATTTGCAAGCGCTACTATATTCGAGCGCGGCTTGACGACGCTGTC 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 LYSLEUVALGINASPRGLYILEASPRVALTHRVALLEUTHRASNUSERLEUGINALATHRASP 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 540 TCAGCCAGGAAAGGCGCGATGCGTCTGCTCTCGCCG-----ATGAACCCGAT 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 376 VALALALAYALHISSEGLYTYRVALLYSTYRGLYSERLEULEUYSALAGLYILE 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 594 CATCCGCTTCTCAAGAGGCGCGCTACACGACCTTCCACGCCCTTTAAAGCGCGCTGC 653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 396 LYSLEUTYRGLILEUGINPROASNHISALAYALPROALATHRYSASPRLYSGLYLEUTHR 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 654 TATATTTCGCGTACTAC-----AGAGCGCTT--- 680
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 416 GLYSERVALTHRSELEUHNISALALYSTHRPHEILEVALSPGLYLSARGILEPHE 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 681 -----TATCATGCAAAAGCATTTAATCTGCATGACCGCATGTGATG 722
QY 436 ILEGLYSERPHEASNLEUASPRPROARGSERVALARGLEUASNTHRGUMETGLYVAL 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 723 ATCGAGACATCGAATTTTGACACAGGAGCGCTGTTCTCAATGATGAGCAATGTCTC 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 456 ILEGLYSERPROLYS-----ILEALGLUGINMETGLUARG 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 783 ATCCATGATTAAGACTGACAAAGCAATTTCTGACGCTGCTCAAGGAACATTTAAAC 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 468 THRLEUALASPHRSEPRROGLUTYRALATYRARGVALTHRLEUASPARGHISASNARG 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 843 GCGAGCTTCTGACAAAGAGCGGTATGCGAAGCGCGGTGATGACGCG-----CCC 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 LEUGINTRP 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 897 GTCGAATGCG 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 39
AAS65339
ID AAS65339 standard; cDNA; 862 BP.
AC AAS65339;
DE 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #1143.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB; ABG01152.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX
XX Claim 1; SEQ ID NO 1143; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human

```



CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 862 BP; 202 A; 212 C; 250 G; 198 T; 0 other;

# Alignment Scores:

Score:	2.16e-08	Length:	862
Percent Similarity:	31.208	Matches:	83
Best Local Similarity:	20.398	Conservative:	44
Query Match:	6.498	Mismatches:	122
		Indels:	158
		Gaps:	12

US-10-066-551-4 (1-525) x AAS65339 (1-862)

```

QY 76 11eTyrlLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSer 95
   ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 ATCCAGTTGCTGGAAACGGCGAGCATATATCCCGCGGTGTTAAGCGCATGCGCAG 105

QY 96 AlaGluHisSerLeuAspLeuGlnIleTyrIleTrrpaGAsnAspIleSerGlyArgLeu 115
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 GCACAGAACGCAATCTTGAACGTTATCTGTTGAGGATGACGTCGCAACAA 165

QY 116 LeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAsp 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 CTGCATGGCGGACACTCTGCGACAGCCGACCGGGGTTAAAGCGAAGTCTTGCTGAT 225

QY 136 AspaAsnArthrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 GCGTACGGTTCGCGGATCTCAGCATGAGTTTGCATGAACTGACGACGTCGCGTA 285

QY 156 GluValArgLeuPheAsnProPheValLeuArgIleTrrpaGAlaLeuGlyTyrIleThr 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 GTGTTCCGCTACTAGATCC-----CGCCCTCGCTTTGTTGATGCGCAC 333

QY 176 AspPheProArgLeuAsnArgArgmethHisAsnIleSerPheThrAlaAspaAsnArgAla 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 AAT-----GTGTTTCGCGGATGATCGCAAAATTTGGTGGATCGACGCGGTATA 384

QY 196 ThrIleLeuGlyGlyArgAsnIleGlyAspGluTyrPheIleValGlyIleAspThrVal 215
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 GCGTTTATTTGGCGGCTAATTACTCCCGCAGCATATGTCAGTACGAGTCCAGAGGCT 444

QY 216 PheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyIleValSerHisAspPhe 235
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 AAACAGATTACGCGGCTACGCGCTTGAAGGCGGATTTGGAAGATATTCTC---CAGTTT 501

QY 236 Asp-----ArgTyrTrpAlaSerHisSerAla 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 GAGCTGGAAACCTCCTCGACAGAGCGCGGACGACGCGCTGTCGACGCTAT----- 555

QY 245 HisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyIleGlnAlaLeuGly 264
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 556 CACAAAGCGGAAAGACAGCCGCGGAGAA-----GCCGACGAGCAACGGA 603

QY 265 TyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSer 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 604 TAT-----GCCGTGTCAAGATTC----- 621

QY 285 ProLeuTyrGlnIleGlnThrArgIleAspTrpGlnSerValGlnThrArgLeu 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 ----- 621

QY 305 IleSerAspAspProAlaAlaGlyLeuAspArgAspArgIleProIleAlaGly 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 -----GCT 624

QY 325 ArgLeuGlnAspAlaLeuGlnProGluLysSerValTyrLeuValSerProTyrPhe 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 625 CGCCTT----- 630
  
```

```

QY 345 ValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 630 ----- 630

QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrVal 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 -----TCTGATATCAT 642

QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHis 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 -----CTGTTAAAGCGCGGTTCAAGGTTTTCAGTACCGCCGCG--- 684

QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 -----CCGCTCCAGCGC 696

QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 697 AAAGTGCATTTGATGACGATCATGCGCAGACAGTAGGGTCCAGTAATCTCGATCCGCTC 756

QY 445 SerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluGln 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 757 AGTTTGTCACTGAATCTGAAAGCAAAATGTCATCATCACGAT-----CGTCAT 804

QY 465 MetGluArgThrLeuAlaAsp 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 805 TTTAACAGACGCTCGCGCAT 825

RESULT 40
AAS65338
ID AAS65338 standard; cDNA; 3294 BP.
AC AAS65338;
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #1142.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
   food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
PF 30-MAR-2001; 2001WO-US08631.
PR 31-MAR-2000; 2000US-0540217.
PP 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
PI MPI: 2001-639362/73.
PI P-PSDB: ABG01151.
PT New isolated polynucleotide and encoded polypeptides, useful in
   diagnostics, forensics, gene mapping, identification of mutations
   responsible for genetic disorders or other traits and to assess
   biodiversity.
PS Claim 1: SEQ ID NO 1142; 103pp: English.
XX The invention relates to isolated polynucleotide (I) and
   polypeptide (II) sequences. (I) is useful as hybridisation probes,
   CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
   CC and gene mapping, and in recombinant production of (II). The
   CC polynucleotides are also used in diagnostics as expressed sequence tags
   CC for identifying expressed genes. (I) is useful in gene therapy techniques
   CC to restore normal activity of (II) or to treat disease states involving
  
```

CC (11) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (11) and its binding partners are useful in medical  
 CC imaging of sites expressing (11). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPRO  
 CC at ftp.wipro.int/pub/published\_pct\_sequences.

XX Sequence 3294 BP: 814 A; 735 C; 865 G; 880 T; 0 other;

# Alignment Scores:

Pred. No.:	1.55e-07	Length:	3294
Score:	176.00	Matches:	83
Percent Similarity:	31.20%	Conservative:	44
Best Local Similarity:	20.39%	Mismatches:	122
Query Match:	6.49%	Indels:	158
DB:	23	Gaps:	12

US-10-066-551-4 (1-525) x AAS65338 (1-3294)

QY 76 ILETyrLeuLeuAsnAspProHisGluAlaIleuAlaAlaAlaLeuIleGluSer 95  
 DB 721 ATCCAGTTGCTGGAAACGGCGAGCAATATATCCCGGGGTTTAAAGCGATTGGCCAG 780  
 QY 96 AlaIleuHisSerLeuAspLeuGlnTyrTyrIleTyrPheAsnAspIleSerGlyArgLeu 115  
 DB .781 GCACAAACACCCATCTTCTGAAACGGTTTATCTGTTTGAAGTGCAGTCGGCAACAA 840  
 QY 116 LeuPheAsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAsp 135  
 DB 841 CTGCATCGCGACTACCTACGCGACGACGACGCGGTTAAAGCGAAGTCTTGCTGGAT 900  
 QY 136 AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle 155  
 DB 901 GGCTACGGTTGCGCGGATCTCAGCATGATGTTTCATGAACTGACGCGCAGCTGGCGTA 960  
 QY 156 GluValArgLeuPheAsnProPheValLeuArgGlyTyrPheAlaLeuGlyTyrLeuThr 175  
 DB 961 GTGTTCCGCTACTACGATCCC-----CGCCCTCGCTTTTGTGATGCGGCACC 1008  
 QY 176 AspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195  
 DB 1009 AAT-----GTGTTTCCCGCGATGATGCANAATTTGTGTGATGACCGCGGTATA 1059  
 QY 196 ThrIleuGlyGlyArgAsnIleGlyAspGlyTyrPheLysValGlyGluAspThrVal 215  
 DB 1060 GCCTTTATTTGGCGGCGTAATTAATCTCCGCGCATATATGTCAGTACGTCGAGAGCT 1119  
 QY 216 PheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAspPhe 235  
 DB 1120 AATCAGATTAACGCGTACGCTTGAAGGCGCATTTGCGAAGATTTCTC---CAGTTT 1176  
 QY 236 Asp-----ArgTyrTrpAlaSerHisSerAla 244  
 DB 1177 GAGCTGGAACCTGCTGCGACAGCGCGCGACGACGCTGTGCGACGTCAT----- 1230  
 QY 245 HisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGlyLeuGlnAlaLeuGly 264  
 DB 1231 CACAAACGCGAAGAGAACCGCGACCGGAGAA-----GCCGACGCGAAGCGA 1278  
 QY 265 TyrAsnAspGlyThrSerArgHisAlaLeuLeuArgGlyTyrArgGlyThrValGluGlnSer 284  
 DB 1279 TAT-----GCCCTGTCTCAGATTTC----- 1296  
 QY 285 ProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeu 304

DB 1296 ----- 1296  
 QY 305 IlesAspAspProAlaLysGlyLeuAspArgAspArgArgProProIleAlaGly 324  
 DB 1297 -----GGT 1299  
 QY 325 ArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValTyrLeuValSerProTyrPhe 344  
 DB 1300 CGCCTT----- 1305  
 QY 345 ValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspVal 364  
 DB 1305 ----- 1305  
 QY 365 ThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrVal 384  
 DB 1306 -----TCTGATATCAT 1317  
 QY 385 LysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHis 404  
 DB 1318 -----CTGGTTAAAGCGCGCGCTCAGGTTTGTGAGTACCGCGCGC--- 1359  
 QY 405 AlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAla 424  
 DB 1360 -----CCGCTCCAGGC 1371  
 QY 425 LysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444  
 DB 1372 AAGTGGCATTTGATGAGCATCTGCGCGACAGTAAAGGTCAGTAATCTGATCCGCTC 1431  
 QY 445 SerAlaArgLeuAsnThrGlnMetGlyValValIleGlnSerProLysIleAlaGluGln 464  
 DB 1432 AGTTGTCACAGAAATCTCGAAGCAAAATGTCAATCAACAGAT-----CGTCAT 1479  
 QY 465 MetGluArgThrLeuAlaAsp 471  
 DB 1480 TTTAACCGACGCTGCGCGAT 1500

Search completed: May 19, 2003, 18:56:24  
 Job time : 6959 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 17:01:11 ; Search time 52 Seconds  
(without alignments)  
3096.256 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 2713  
Sequence: 1 MRANPKTQAMPSETISLMKT.....KLWKRIAKILSLPIESLL 525

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 segs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=x1h  
-O=/cgn2.1/USPTO.spool/US10066551/rnat.12052003.091143.22378/app\_query.fasta.1.711  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=tni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=numan40.cdi  
-LIST=45 -DOCALLIN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=40  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US10066551.@CGN.1.1.32.@unat.12052003.091143.22378 -NCPU=6 -ICPU=3  
-NO.XLPXY -NO.MMAP -LARGEJURY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA.\*  
1: /cgn2.6/ptodata/1/ina/5A.COMB.seq.\*  
2: /cgn2.6/ptodata/1/ina/5B.COMB.seq.\*  
3: /cgn2.6/ptodata/1/ina/6A.COMB.seq.\*  
4: /cgn2.6/ptodata/1/ina/6B.COMB.seq.\*  
5: /cgn2.6/ptodata/1/ina/PCTUS.COMB.seq.\*  
6: /cgn2.6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	295	10.9	1488	4	US-09-134-001C-1436 Sequence 1436, Ap
2	264.5	9.7	1485	4	US-09-134-001C-1339 Sequence 1339, Ap
3	253	9.3	2234	4	US-09-221-017B-957 Sequence 957, App
4	145.5	5.4	3388	4	US-09-141-206-1 Sequence 1, Appl
5	133.5	4.8	3425	4	US-09-107-149-16 Sequence 16, Appl
6	130.5	4.8	2799	4	US-08-968-752B-5 Sequence 5, Appl
7	130.5	4.8	2799	4	US-09-536-224-5 Sequence 5, Appl
8	121.5	4.5	3425	2	US-08-768-147B-1 Sequence 1, Appl
9	121.5	4.5	3425	2	US-09-107-149-1 Sequence 1, Appl
10	113.5	4.2	13926	4	US-08-961-527-5 Sequence 5, Appl
11	111	4.1	3108	3	US-08-968-752B-3 Sequence 3, Appl
12	111	4.1	3108	4	US-09-536-224-3 Sequence 3, Appl

13	106	3.9	2502	1	US-08-073-384C-21 Sequence 21, Appl
14	106	3.9	2502	1	US-08-254-359A-21 Sequence 21, Appl
15	106	3.9	2502	1	US-08-483-043-21 Sequence 21, Appl
16	106	3.9	2502	1	US-08-481-238-21 Sequence 21, Appl
17	106	3.9	2502	2	US-08-471-066B-21 Sequence 21, Appl
18	106	3.9	2502	2	US-08-484-956-21 Sequence 21, Appl
19	106	3.9	2502	2	US-08-757-653-21 Sequence 21, Appl
20	106	3.9	2502	2	US-08-599-491-21 Sequence 21, Appl
21	106	3.9	2502	2	US-08-756-386-21 Sequence 21, Appl
22	106	3.9	2502	2	US-08-823-516-21 Sequence 21, Appl
23	106	3.9	2502	3	US-08-682-853A-21 Sequence 21, Appl
24	106	3.9	2502	3	US-08-759-038-21 Sequence 21, Appl
25	106	3.9	2502	3	US-08-758-314-21 Sequence 21, Appl
26	106	3.9	2502	4	US-09-350-309-21 Sequence 21, Appl
27	106	3.9	2502	4	US-08-520-946-21 Sequence 21, Appl
28	106	3.9	2502	2	US-08-757-653-187 Sequence 187, App
29	106	3.9	2505	2	US-08-823-516-68 Sequence 68, Appl
30	106	3.9	2505	3	US-08-759-038-129 Sequence 129, App
31	106	3.9	2505	3	US-08-758-314-129 Sequence 129, App
32	105.5	3.9	28804	2	US-08-592-874-1 Sequence 1, Appl
33	105.5	3.9	28804	3	US-09-096-942-2 Sequence 2, Appl
34	105.5	3.9	28804	3	US-09-096-867-2 Sequence 2, Appl
35	105	3.9	2505	2	US-08-757-653-167 Sequence 167, App
36	105	3.9	2505	2	US-08-757-653-189 Sequence 189, App
37	105	3.9	2505	2	US-08-823-516-65 Sequence 65, Appl
38	105	3.9	2505	2	US-08-823-516-70 Sequence 70, Appl
39	105	3.9	2505	3	US-08-759-038-106 Sequence 106, App
40	105	3.9	2505	3	US-08-759-038-131 Sequence 131, App
41	105	3.9	2505	3	US-08-758-314-106 Sequence 106, App
42	105	3.9	2505	3	US-08-758-314-131 Sequence 131, App
43	104.5	3.9	9686	4	US-09-221-017B-1003 Sequence 1003, App
44	103.5	3.8	5305	4	US-08-961-527-135 Sequence 135, App
45	103	3.8	2693	1	US-07-872-644-50 Sequence 50, Appl

## ALIGNMENTS

RESULT 1  
US-09-134-001C-1436  
Sequence 1436, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC  
FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GFC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 1436  
LENGTH: 1488  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-1436

Alignment Scores:  
Pred. No.: 6.8e-25  
Score: 295.00  
Percent Similarity: 38.85%  
Best Local Similarity: 23.68%  
Query Match: 10.87%  
DB: 4  
Gaps: 14

US-10-066-551-4 (1-525) x US-09-134-001C-1436 (1-1488)  
Oy 74 SeraspIeTyRleuLeuAsnAspProHisGuaIaPheAlaIaArgAlaLeuIle 93  
Db 394 AATGATCTACAAATTATTACAGATGTCATCAGAAAGTTCGATGATTATATGATATT 453

OY	94	GIUSERIALAGLUNHSSErSLeuAspLeuSrgLlyrTyrLlePrtrpArgAsnAPrLleSergLY	113
		: : : : :	
DB	454	AGACACGCACATCATCATATATTCACATCCATACACTACCTATTATCTACATGTAAATCTCGGA	513
OY	114	ArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluAArgLysValArgValArgLeuLeu	1133
		: : : : :             : : : : :	
DB	514	AAGCAATGTGTTACATGATTTAGAAAAAAACCTGAAAGAGTATAGAAAGTTAAAGTCA	573
OY	134	LeuAspAspAsnAsrThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisPro	153
		: : : : :	
DB	574	TACGATGTGATGTGGCTCAAGAGATTTA-----	600
OY	154	AsnLleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyr	173
		: : : : :	
DB	601	-----CGAAAGAAAGTCTAAAAAAAGTTTCTCAAAAAGCGCG	639
OY	174	LeuThrAsp-----PhePro-----	183
		: : : : : ArgLeuAsnArgArg	
DB	640	CACCGCTGACTTTCTTCTTCATCCATCAAAATTACCACTATTATTAATAAGAAATTAATAGA	699
OY	184	MetHisAsnLysSerPheThrAlaAspAsnArgAlaThrLleLeuGlyArgTyrAsnIle	203
		: : : : :	
DB	700	AACGCACAAAAAATAGTAGTCATTGACCGCTACATCGATATGTGGTGGTTTAATGTA	759
OY	204	GlyAspGlyTyrPheLysValGlyGluAspThrValPheAla-----AspLeuAsp	220
		: : : : :	
DB	760	GGAGATGTGATAT-----ATAGGTAAATCTAGAAATTTGGTATTGGAGGACACACAC	813
OY	221	LleLeuAlaThrGlySerValValGlyValSerHisAspPheAspArgTyrTTPAla	240
DB	814	TTACGATTTAAAGGTGACCGCGGTAAATGCCCTTCACTGATTCATCTTAGATTGGAC	873
OY	241	SerHisSerAlaHis--AsnAlaThrArgLleLeuArg-----	252
DB	874	TCACATCAACTCGCGACAAATTTACTTATTAATCGATATTTCTCGATGTGACTCT	933
OY	253	SerGlyAsnLleGlyLysGlyLeuGlnAlaLeuGlyLysAsnArgLysThrSerArgHis	272
		: : : : :       : : : : :	
DB	934	GGTGTACGATAGGCATACCAAAATGTCTCAAGTGGCGCTAAGAA-----	978
OY	273	AlaLeuLeuArgTyrArgGlyThrValGlnLleSerProLeuTyrGlnLysLleGlnThr	292
		-----	
DB	978	-----	978
OY	293	GlyArgLleAspTrpGlnSerValGlnThrArgLeuLleSerAspAspProAlaLysGly	312
		: : : : :	
DB	979	-----GATGTGGAAACAATTAAG-----	996
OY	313	LeuAspArgAspArgArgLysProProLleAlaGlyArgLeuGlnAspAlaLeuLysGln	332
DB	997	-----TATCGTTATTTTAAAG--ATGATTTTCATCA	102
OY	333	ProGlyLysSerValTyrLeuValSerProTyrGlyPheValProThrLysSergLysThrAsp	352
		: : : : :	
DB	1024	GCCAAAGATCATATATCATTCATCAACATCACTCACTTCATTCGCCATCCAAAGCTTTTATGAT	108
OY	353	AlaLeuAlaLysLeuValGlnAspGlyLleAspValThrValLeuThrHisSerLeuGln	372
		: : : : :	
DB	1084	TGCATTTAAATTCGACCGCTTGGTGGTGGTGGTGGTATACATCATAGTGTCTTAACAAACG	114
OY	373	AlaThrAspValAlaAlaValaHisSerGlyTyrValLysTyrArgLysProLeuLeuLys	392
		: : : : :	
DB	1144	-----GATCATCCCTTTGTTATTTGGGCTACGCTTAAAAACGTAAGCGTTCATTGGA	119
OY	393	AlaGlyLleLysLeuTyrGlnLeuGlnProAsnHisAlaValProAlaThrLysAspLys	412
		: : : : :	
DB	1198	GCTGCGCTTATGTCTATCATTTAT-----GATTAAT	122
OY	413	GlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheLleValAspGlyLys	432
DB	1228	GGTTTT-----CTACATTCAAAACACTGTGTGATTCATGATGAG	126
OY	433	ArgLlePheLleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMet	452

Db	1267	GTAGCAAGTGTAGGACACTGCCAATATGCAATATGCAAGCTTACTTAACTTTAGAGTCC	1326
Oy	453	GIYValIleGluSerProLysIleAlaGluGluMetGluArgThrLeuAlaAsp---	471
Db	1327	AATGCTTTTATTTATGATGAAGCGCTACACAGATCATTTAAACAGCTTTCATTAATGAT	1386
Oy	472	-----ThrSerProGluTyrAlaTyrArg	479
Db	1387	ATGAATAATTACTACACAGCTCACTTGTGAAGAATATTTGTTAAACGA	1431
RESULT 2			
US-09-134-001C-1339			
; Sequence 1339, Application US/09134001C			
; Patent No. 6380370			
; GENERAL INFORMATION:			
; APPLICANT: Lym Doucette-Stamm et al			
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC			
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS			
; FILE REFERENCE: GTC-007			
; CURRENT APPLICATION NUMBER: US/09/134,001C			
; CURRENT FILING DATE: 1998-08-13			
; PRIOR APPLICATION NUMBER: US 60/064,964			
; PRIOR FILING DATE: 1997-11-06			
; PRIOR APPLICATION NUMBER: US 60/055,779			
; PRIOR FILING DATE: 1997-08-14			
; NUMBER OF SEQ ID NOS: 5674			
; SEQ ID NO 1339			
; LENGTH: 1485			
; TYPE: DNA			
; ORGANISM: Staphylococcus epidermidis			
US-09-134-001C-1339			
;			
Alignment Scores:			
Pred. No.:	2,79e-21	Length:	1485
Score:	264.50	Matches:	101
Percent Similarity:	38.05%	Conservative:	71
Best Local Similarity:	22.35%	Mismatches:	177
Query Match:	9.75%	Indels:	103
Db:	4	Gaps:	11
US-10-066-551-4 (1-525) x US-09-134-001C-1339 (1-1485)			
Oy	76	IleTyrIleuLeuAsnAspProHisGluAlaIlePheAlaIleAlaLeuIleGluSer	95
Db	397	ATCGATTATATTACAGATGGTCATTAAGCTTTATGAAAAAGTACTTGAGCATTTTACAT	456
Oy	96	AlaGluHisSerLeuAspLeuGluGlnTyrTyrIleThrPArgAsnAspDileSerGlyArgLeu	115
Db	457	GCTCAAGACTATATACATCTAGACTACTATACCTTTGAACCTGANTGATTAAGTAAAGA	516
Oy	116	LeuPheAsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAsp	135
Db	517	ATCTTAGTAGTCACCTTGAAGACTTAACCTTAAAGAGCTTTAGAAGTTAAACTTTGTATGAC	576
Oy	136	AspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIle	155
Db	577	GATGTGGTCTAAA-----	591
Oy	156	GluValArgLeuPheAsnProPheValLeuArgTyrStrPArgAlaLeuGlyTyrLeuThr	175
Db	592	AAGGTATAGATTATCAAAA-----TTTAAACATTTTCAGAGCAWTAGTGGAGAACTT	642
Oy	176	Asp-----PhePro-----ArgLeuAsnArgArgMetHis	185
Db	643	GAAAGCATTTTCCCTCGAAGAGTACTTTAATCAATTTTCAGATGAATAATCGAAATCAT	702
Oy	186	AsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGluGlyArgAsnIleGlyAsp	205
Db	703	AGAAAGATTATCATATATAGATGAGACAATTTGGTTACCTTGGCGGTTTAAATGTCGGCAT	762
Oy	206	GluTyrPheIleValGlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGly	225

```

Db 763 GATTATTAGGATTAGTAAGTGTACTGAGAGATACATACAGCTGTTCACAGT 822
Qy 226 SerValValGlyGluValSerHisAspPheAspArgTyrThrAlaSerHisSerAlaHis 245
Db 823 GAATGCATCGACGACATACATTAAGATTATTTAGACTGCAATTCACAGTCG----- 876
Qy 246 AsnAlaThrArgIleIleArgSerGlyAsnIleGlySerGlyLeuGlnAlaLeuGlyTyr 265
Db 876 ----- 876
Qy 266 AsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerPro 285
Db 877 -----CATGCTCCACAATTAAATTGATCA 903
Qy 286 LeuTyrGlnIleGlnThrGlyArgIleAspThrGlnSerValGlnThrArgLeuLe 305
Db 904 AAATATTTCCTTAAAAAATGG-----GACAAAGGAACGGCGCTATTCAATCGCT 957
Qy 306 SerAspAspProAlaIleGlyLeuAspArgAspArgGlySerProIleAlaGlyArg 325
Db 958 TCTAGTGACCTGCATTGTGATTACATCA-----ATGAATATGCT 999
Qy 326 LeuGlnAspAlaLeuIleGlnProGluIleSerValTyrLeuValSerProTyrPheVal 345
Db 1000 TATACAAAATGATGATGAGCGCTAAAAAGCTATCTATCTACAAAGCCCTTACTTAT 1059
Qy 346 ProThrIleSerGlyThrAspAlaLeuAlaIleValGlnAspGlyIleAspValThr 365
Db 1060 CCAAGCCATATACATTAATGATTAATAAATGGCTGTAATAGCGCGGTGGAAGTAAC 1119
Qy 366 ValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValIys 385
Db 1120 CTATATGATACCGTGT-----AAACGTCATCATCATCGTTTATTGGGTACATTTCA 1173
Qy 386 TyrArgIleProLeuLeuIleGlyIleLeuLeuTyrGluLeuGlnProAsnHisAla 405
Db 1174 AATCAGCTGATTTATGATGATGAGGAGTATATTTACCTTATCAATAATGATTT-- 1230
Qy 406 ValProAlaThrIleAspIleGlyLeuThrIleSerSerValThrSerLeuHisAlaIys 425
Db 1231 -----ATTCATTCCTAAA 1242
Qy 426 ThrPheIleValAspGlyIleAspIlePheIleGlySerPheAsnLeuAspProArgSer 445
Db 1243 ATATTAAATGATGATGATGATGATTTCTTCATTTGGTGTAGTGAACATGAGCTTTAGAGC 1302
Qy 446 AlaArgLeuAsnThrGlnIleGlyValValIleGluSerProIleIleAlaGluGlnMet 465
Db 1303 TTGGAAGCTGATTCGAAAGTGAATGCATTTATATATGATGAAGATATAGCTAAACAGTTA 1362
Qy 466 GluArgThrIleValAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHis 485
Db 1363 CGTCAA-----GCATTTGAAAAAGATATGAGCAATCT 1395
Qy 486 AsnArgLeuGlnThrPheAspProAlaThrArgIleGlyThrTyrProAsnGluProGluAla 505
Db 1396 AAATTACTT-----ACAAAGAGATATAGCAACCGCTCATATCA 1437
Qy 506 LysLeuThrIleAlaAlaAlaLysIleLeuSer 517
Db 1438 ATTAATTTAAAGAGACTAGCAAAATTAATATCT 1473

```

## RESULT 3.

```

US-09-221-017B-957
Sequence 957, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD

```

```

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 957:
SEQUENCE CHARACTERISTICS:
LENGTH: 2234 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
AMT-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...2234
US-09-221-017B-957
Alignment Scores:
Pred. No.: 1.28e-19 Length: 2234
Score: 253.00 Matches: 101
Percent Similarity: 38.33% Conservative: 73
Best Local Similarity: 22.25% Mismatches: 142
Query Match: 9.33% Indels: 138
DB: 4 Gaps: 16
US-10-066-551-4 (1-525) x US-09-221-017B-957 (1-2234)
Qy 49 HisPheAsnThrSerIleProValArgLeuAspAsnIleLeuGlnIleArgHisThrPro 68
Db 478 CACTATCTGTCATCCCGGACGACAGTGTGAAGCATGACATATATCCAGCTGAATCCC 537
Qy 69 -----HisThrAsnGlyLeuSerAspIleTyrLeuLeuAsnAspProHis 83
Db 538 ATCATGAAGCTGATCAACACACATCGGACACCCCGTC--ATGCTCAGCTCGCCCGCG 594
Qy 84 GluAlaPheAlaIleArgAlaAlaLeu-----IleGluSerAla 96
Db 595 CATATATTCACGCGACGAGAGATATGTTTCGACTGCTTCTTCGGACATATAGAGCTGCC 654
Qy 97 GluHisSerLeuAspLeuGlnIleTyrTyrIleTyrPargAsnAspIleSerGlyArgLeuLeu 116

```

```

Db 655 GCCACCATGTCATATTCACTTACGTCAATGACATGACATCTGGGACATGCCCTG 714
QY 117 PheasnleuValTyrleuAlaIleuArgGlyValArgValArgleuLeuLeuAsp 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 TCCGATCTGCTTATACGACAGGTGACAGGCGGTGAAGGTGAGGCTCATTTGCTCAT 774
QY 137 AsnAsnThr-----ArgGlyLeuAspAspleu-----LeuAlaLeuAsp 150
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 775 GTGGGTTCTGGGCTACGGCGGTGGCTACTGGAACGACGCGCGGAGGTGAA 834
QY 151 SerHisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLysThrArgAla 170
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 835 GCCATCTCTTATGCGCGTC-----TTCTCCCATACG----- 870
QY 171 LeuGlyTyrleuThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThr 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 871 -----TCCAGCGCGGTAAACTACCGCATACCGGAGGTAGCGCTT 912
QY 191 AlaAspAsnArgAlaThrIleLeuGlyArgAsnIleGlyAspGluTyrPheLysVal 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 913 ATGATGTAATAATAGCTATTTGGCGGTATGATGATGGCGCATCTTACTCGGCG 972
QY 211 GlyLysAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValGlyGlu 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 973 AACAGGCTCGGCTGTGGCGGATACCTCGCTATGATGAGGCGCATGTGCGCGGT 1032
QY 231 ValSerHisAspPhe-----AspArgTyrThrAlaSerHisSerAlaHisAsnAlaThr 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1033 TTGAGAGCAGTTTCTGCTGGACTGGTATGTCGTATCGCACGAGTCAATCTCTCC 1092
QY 249 -----ArgIleIle 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1093 GCGTACTATCCGATCTGCCGATACCCGATCCGATCCGACGAGGTGGAAGATACACTGTA 1152
QY 252 ArgSerGlyAsnIleGlyGlyGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArg 271
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1153 CGCTCCGCGCCCGTGGGCAATG----- 1176
QY 272 HisAlaLeuLeuArgTyrArgGluThrValGlnGlnSerProLeuTyrGlnLysIleGln 291
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1177 -----CGTACCATGACGAGCGCCACCATCTAC----- 1203
QY 292 ThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAspProAlaLys 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1203 ----- 1203
QY 312 GlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLys 331
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 -----TGTATCGAG 1212
QY 332 GlnProGlyLysSerValTyrleuValSerProTrpPheValProThrLysSerGlyThr 351
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1213 CGTGCCGCAACATGCTATATGACAGCTCTTACTCTCCGACCGAAGAGCTGAAT 1272
QY 352 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThr----- 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1273 AATGCTCTCATACGAGAGCTTTGGCGGCGCATGATCTACACCTGCGTGGCCCAACGA 1332
QY 369 AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLys 388
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1333 AGCGATGCCCGCTGCGCACAGTGGCGGC-----TTCTCTATGTGGAG 1377
QY 389 ProLeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaValProAla 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1378 GATCTCTTGAGGCGCGGTGATCATGTACAC-----CTCTATCCAAC 1419
QY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIle 428
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 -----GGCTTTCTCCATTTCCAAAGCTGATATG 1446
QY 429 ValAspGlyLysArgIlePheIleGlySerPheAsnLeuAsnProArgSerAlaArgLeu 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1447 GTGGAATAGTCAGTGCTGTGATCGGCTTACGAATATGACATCTCCGACGCTTGGAAT 1506

```

```

QY 449 AsnThrGlnMet-----GlyValValIleGluSerProLysIle 461
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1507 AACCTTCGATTCATTCATGATATCATATGACAGAGCTACTGTC 1548

RESULT 4
US-09-141-206-1
: Sequence 1, Application US/09141206
: Patent No. 6187559
: GENERAL INFORMATION:
: APPLICANT: Steed, Paul M.
: APPLICANT: Lasala, Daniel J.
: TITLE OF INVENTION: Amino Acid Sequence of Human PLD2A
: Patent No. 6187559
: FILE REFERENCE: 4-30148/P1/CGC1954/R
: CURRENT APPLICATION NUMBER: US/09/141,206
: EARLIER FILING DATE: 1998-08-27
: EARLIER APPLICATION NUMBER: 60/057,802
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3388
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (162)..(2963)
: OTHER INFORMATION: Human PLD2
US-09-141-206-1

Alignment Scores:
Pred. No.: 1.43e-06 Length: 3388
Score: 145.50 Matches: 123
Percent Similarity: 32.06% Conservative: 87
Best Local Similarity: 18.78% Mismatches: 180
Query Match: 5.36% Indels: 265
DB: 4 Gaps: 32

US-10-066-551-4 (1-525) x US-09-141-206-1 (1-3388)
QY 30 LeuLeuCySerCySerSer-----ThrLeuProProLeuGluGlu 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1026 TTGATTTCTCAAGTGCACACAGCTACCGGACGACGAGGTGGGCCCAAGAAATCACTGAG 1085
QY 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1086 CTGGCACAGGCG-----CCAGGCAGA-----AACTTCCTACAG 1118
QY 64 Ile---ArgHis-----ThrProHisThrAsnGlyLeuSerAspIleTyrLeu 78
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1119 CTGACCGGATGACACACTACGCCGCCGCCGCTGGGAAGTGGCCGCTGTTGTG 1178
QY 79 LeuAsnAspProHisGluAlaPheAlaAlaArgAlaIleLeu----- 92
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 AATGGCGCAAGTACTTCTGCTGCTGGCAGATGCCATCTCTGAGCTCAAGAGAGATT 1238
QY 93 -----IleGluSerAlaGluHisSerLeu 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1239 TTCATCACAGACTGGTGTGAGTCTGAGGTTTACTGAAGCGTCCGCGCATTTGAT 1298
QY 101 AspLeuGlnTyrTyrIleTTPArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuVal 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1299 GAC-----TGAGACGTGACATT-----ATGTTTC 1322
QY 121 TyrleuAlaIleGluArgGlyValArgValArgLeuLeuLeuAspAsnAsnThrArg 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1323 AAGAGAGAGCGGAGAGAGGTGTCGTTGTTCTATTTCTGTTTAAGAA----- 1373
QY 141 GlyLeuAspAspLeuLeuLeuAlaLeuAspSer----- 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1374 -----GTGAATTTGGCTTGGGCGATCAACAGTGGCTATGCAAGAGGCGGTGATGCTG 1427

```

OY	152	----	HisProAsnIleGluValaLysLeuPheAsnProPheValLeuArgLysTrpArgAla	170
Db	1428	CTGCACCCCAACATMAAGGNG	-----ATGCGTCACCGACAGCAAGTAGACGTGGG-----	1478
OY	171	LeuGlyTrpLeuThrAspPheProArgLeuAsnArgMetHisAsnLysSerPheThr	190	
Db	1479	-----	-----GCCCATCATGAGAGAGCTCTGGTG	1502
OY	191	AlaAspAsnArgAlaThrIleLeuGlyGly	-----	200
Db	1503	GTGGACCAAGTGGTACCATTC	CTGGGGGGGACTGGACCTTGCTATAGCCGCTGGGATGAC	1562
OY	201	-----	-----ArgAsnIleGlyAsp	205
Db	1563	CTGCACCTACCGACTGACTGACCTTGGAGACTTTCGAAATAGCTCTCCAGCTTCC	1622	
OY	206	-----	-----GluTrpPheLysVal	210
Db	1623	ACCCCGGGCCACAGCTACCAAGCACCACAGACTTCTTACACACCAATCTTCTGGCTG	1682	
OY	211	GlyGluAsp	-----	213
Db	1683	GGCAAGAGACTACAGCATCTTATCCACCAAGACTGGGTGCTGACCTGGACCGGCTTTCGA	1742	
OY	214	-----	-----ThrValPheAlaAspLeuAspIleLeuAla	223
Db	1743	GATTTCATTGTGACAGGAGACGACCCCTCGGATGCCATGGCGGGAGCTGGGTGGTCTCTC	1802	
OY	224	ThrGlySerValValGlyGluValSerHisAspPheAspArgTrpAlaSerHisSer	243	
Db	1803	CATGGCTACCGCGCGGGACCTTGGCCCGGACCTTATCCACGCTGG	-----	1850
OY	244	AlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeu	263	
Db	1851	-----AATTTCACCAAGACACCAAGGCGC	-----	1874
OY	264	GlyTrpAsnAspGluThrSerArgHisAlaLeuLeuAlaTrpArgGluThrValGluGln	283	
Db	1875	AAGTACAAGATCCCACTACATCCCTCCCTGCTCCCAACCTTTACACAGCGCCGCAATGAG	1934	
OY	284	SerProLeuTrpGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnPheArg	303	
Db	1935	TTCCCTTC-----ACACTTCCAGAGGGGCACTGC	-----ACCACTCTACAGCTC	1979
OY	304	LeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysProIleAla	323	
Db	1980	-----TTGGCATCATGAGGACCGCTGTCA	-----GCA	2006
OY	324	GlyArgLeuGlnAspAla	-----LeuLysGlnProGluLys	335
Db	2007	GGGACTCTGGAGAACTCATCTCAATGCTACTGTCACACATCATGAGGAGGACCCAGAC	2066	
OY	336	SerValTrpLeuValSerProTrpPheValProThrLysSerGlyThrAspAlaLeuAla	355	
Db	2067	TTCCCTCATCATGTGAAATCATGTTCTTCATTAAGTCTGCTCAGATGGGGGACGGTTCTTGAC	2128	
OY	356	LysLeuValGlnAspGlyIleAsp	-----	363
Db	2127	AAGGTGGGCGCATGAGATTGGAGACAGAACTCTGAAGCCCAAAACAGGGGGTGGTTTAC	2186	
OY	364	-----ValThrValLeuThr	-----	368
Db	2187	CGAGCTACAGCTGTTTGGCCCTTACTCTCCGCTTCGAGAGGTGACATCTCCACGGCGGT	2246	
OY	369	-----AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTrp	383	
Db	2247	GGCACTTCATCCAG-----GCCATTCTGCACCTTACTTACAGGACCTGTGT	2294	
OY	384	-----ValLysTrpArgLys	388	
Db	2295	CGTGGGAGATTTCATCTCGCATCCGCTTAAAGCAGCCATGGGAGACACATGGCGGAGC	2354	
OY	389	ProLeuLeuLysAlaGlyIleLysLysLeuTrpGluLeuGlnProAsnHisAlaValProAla	408	

```

Db 2355 TATATTTCCATCTGCGGGGCTTGATACACAGGAGAGCTGGGGGACCCCTC----- 111 111
Oy 409 ThrLysaspLysGlyLeuThrGlySerValThrSerLeuHisAlaLysThrPhe 428
Db 2409 -----TCGAGACTCATCTTACATCCACAGAGGCTGTCTATC 2444
Oy 429 ValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSer-----Ala 446
Db 2445 GCACATGACCGGACAGTCATCATCTGGTTGTGGCAACATCAATGACCGGAGCTTGCTGGG 2504
Oy 447 ArgLeuAsnThrGlnMetGlyValValIleGlu-----SerProLysIle-- 461
Db 2505 AAGCGGACAGTGTAGCTGGCGCTGTGATCGAGACACAGACAGACGAGACCATCCCTATG 2564
Oy 462 -----AlaGlnGlnMetGluArgThrLeuAlaAspThrSerProGluThrAlaArg 479
Db 2565 AATGGCGAGAGTATCAGCGGGGACAGTTGCTTCCTTAGCTCGGAAACACATGCTTCGCT 2624
Oy 480 ValThrLeuAspArgHisAsnArg-----LeuGlnThrPheAspProAlaThrArgLys 497
Db 2625 GTGATGTTCTTGAGAACAAATACCGGGCCAGACTTGATCTCCGAGACCCATCTGTGATGAC 2684
Oy 498 ThrLysrProAsnGluProGluAlaLysLeuThrLysArgIleAla 512
Db 2685 TTCTTCTC-----CAGTTGTGGCAAGACATGCT 2711

```

```

RESULT 5
US-09-107-149-16
; Sequence 16, Application US/09107149
; Patent No. 6274363
; GENERAL INFORMATION:
; APPLICANT: Leung, David W.
; APPLICANT: Tompkins, Christopher K.
; TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
; FILE REFERENCE: 077319/014
; CURRENT APPLICATION NUMBER: US/09/107,149
; CURRENT FILING DATE: 1998-06-30
; EARLIER APPLICATION NUMBER: 08/768,147
; EARLIER FILING DATE: 1996-12-17
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 3425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64)..(2862)
US-09-107-149-16

```

Alignment Scores:	
Pred. No.:	3.84e-05
Score:	133.50
Percent Similarity:	32.00%
Best Local Similarity:	18.15%
Query Match:	4.92%
DB:	4
Gaps:	35
Length:	342
Matches:	118
Conservative:	90
Mismatches:	187
Indels:	31
Gaps:	25

US-10-066-551-4 (1-525) x US-09-107-149-16 (1-3425)

Qy 30 LeuLeuCySerCySerSer-----TrrLeuProProLeuGlu 43  
|||::: ||||| ||| |||  
Db 928 TTTGATTCTCAAGTCCACACCTACCGGACGACGGTGTGGCCCAAGATCTCTG 987  
Qy 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeu 63  
|||::: ||| ||| |||  
Db 988 CTGGCACAGGGC-----CCAGGCAGA-----GACTTCTCTACG 1020  
Qy 64 Ile---ArgHis-----ThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeu 79  
|||::: ||| ||| |||  
Db 1021 CTGACCGGGCATGACAGCTACGGCCCAACCCCGGCTGTGGACCTTG6CCGGGTGGTTGTG 1080







GENERAL INFORMATION:  
APPLICANT: Frohman, Michael A.  
APPLICANT: Morris, Andrew  
TITLE OF INVENTION: No. 637965e1 Phospholipase D Polypeptide and  
TITLE OF INVENTION: DNA Sequences  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ONYX Pharmaceuticals, Inc.  
STREET: 3031 Research Drive  
CITY: Richmond  
STATE: California  
COUNTRY: USA  
ZIP: 94806  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/536,224  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/968,752  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Giotta, Gregory J.  
REGISTRATION NUMBER: 32,028  
REFERENCE/DOCKET NUMBER: ONYX2004  
TELEPHONE: 510-222-9700  
TELEFAX: 510-222-9758  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2799 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2799  
US-09-536-224-5  
Alignment Scores:  
Pred. No.: 6,19e-05 length: 2799  
Score: 130.50 Matches: 99  
Percent Similarity: 33.77% Conservative: 82  
Best Local Similarity: 18.47% Mismatches: 172  
Query Match: 4.81% Indels: 183  
Gaps: 24  
US-10-066-551-4 (1-525) x US-09-536-224-5 (1-2799)  
QY 98 HisSerLeuAspLeuGlnTyrTrpArgAsnAspIleSerGlyArgLeuLeuPhe 117  
Db 1129 CATCCGACGAC-----TGAGACTGCACATT-----1155  
QY 118 AsnLeuValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAspAsn 137  
Db 1156 ---ATGCTCAAGAGAGGCGGAGAGGTGTCGAGTTTCCATVACGTGTTTAAAGAA 1212  
QY 138 AsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer-----151  
Db 1213 -----GTGAGAGCTGCGCTTGCGCATCAACAGTCTCAGCAAGAGAGC 1257  
QY 152 -----HisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLys 167  
Db 1258 CTGATGCTGCTGATCCACATTAAGGTG---ATGCGACACCCAGACCTTGTCACACTG 1314  
QY 168 TTrpArgAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgMetHisAsnLys 187

Db 1315 TGG-----111  
QY 188 SerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGly-----204  
Db 1333 CTCCTGCTGTGAGACCAAGTGTGGCATTTCTTGGCGGGCTGGACCTGCGCTTGGCCGC 1392  
QY 205 -----AspGlnTyrPheLys-----ValGlyGluAspThrValPheAlaAspLeu 219  
Db 1393 TGGGATGACGTCGAATACCGACCTGACTGACCTGGTGACCCCTGTGAACCTGTACATTTA 1452  
QY 220 AspIleLeuAlaThrGlySer-----ValValGlyCysThrValSerHisAsp-----234  
Db 1453 CAGACTCCACACTAGTGTACAGCCCTGCAGCCACTCCAGACCTTCGCACTAACAATTTC 1512  
QY 235 -----PheAspArg 237  
Db 1513 TTCCTGCGGAAAGACTACAGCACTCATACCAAGACCTGGGTGCAGCTGGACCGG 1572  
QY 238 TyrTrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGly 257  
Db 1573 CCTTTGACAGATTTCATCGACAGGAGACCAACCCAGATGCCATGGAGGATGTGGA 1632  
QY 258 LysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuAlaGlyTyr 277  
Db 1633 GTGGTTGTACACGAGAGTGTGCGAGGAGC---CTTCCCGGCACTTCATCAGCCGTGG 1689  
QY 278 ArgGluThrValGlu-----GlnSerProLeuTyrGlnLysIle-----290  
Db 1690 AATTTCACCAAGACCACCAAGCCAGGTATAGACACTTTGTRACCTTACCTGCTGCC 1749  
QY 291 -----GlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSer 306  
Db 1750 AAGTCACACGACACTGCACAAACATCTCCCTCATGATCCAGCGGGCATGTGCCACT 1809  
QY 307 AspAspProAlaLysGlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeu 326  
Db 1810 GTGCAGGTCTTGAGTGTGTGATCGATGCTCA-----GCAGGACATTG 1854  
QY 327 GlnAspAla-----LeuLysGlnProGluLysSerValTyr 338  
Db 1855 GAGACATCCATCCCATGCTGCTACCTACATACCATTCGAGAGAGCCGACCTTCTCTAC 1914  
QY 339 LeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuVal 358  
Db 1915 ATTGAGATCAGTCTTCTTACCTAGCTGCTCAGATGGCGAAGTTCGACAAAGGTGGCG 1974  
QY 359 GlnAspGlyIleAsp-----ValThr 365  
Db 1975 GATGAGATTGTGGACAGATCTGAAGGCTCAGACAGACGGGCGAGTGTCCGAGTCTAC 2034  
QY 366 ValLeuThr-----AsnSer 370  
Db 2035 TTGCTTCTGCGCTTTCGCCCTGCGCTTGAGGGGAGACATCCACAGGGGTGTGTAACTCC 2094  
QY 371 LeuGlnAlaThrAspValAlaAlaValHisSerGlyTyr-----383  
Db 2095 ATCCAG-----GCTATTCTGCACTTCACCTACAGGAGCCCTGTCTCGGGGAA 2142  
QY 384 -----ValLysTyrArgLysProLeuLeu 391  
Db 2143 CATTCAACTCTACATGCTCTCAAGACGCCATGGGAGCTCGGTGGCGAGATTACATGTC 2202  
QY 392 LysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAsp 411  
Db 2203 ATCTGTGCGCTTGCCACCCATGAGAGCTGGGGGCGACCCATC-----2247  
QY 412 LysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGly 431  
Db 2248 -----TCTGAGCTCATATATATCCAGCAAGATGTCATTCATGGGAGTAC 2292  
QY 432 LysArgIlePheIleGlySerPheAsnLeuAspProArgSer-----AlaArgLeuAsn 449

Db 2293 AGACAGTCATCATGTTCTGCGAACAATCATGACAGAGCTTCTGGGGAACGCTGAC 2352  
 Oy 450 ThrGluMetGlyValIleGluSerProLysIleAlaGluMetGluArgThrLeu 469  
 Db 2353 AGTGAGCTAGCATCTCTGTCAAGAC-----ACGAAATGGAACCATCCCTC 2400  
 Oy 470 AlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArg----- 487  
 Db 2401 ATGATGATGGGTGAGTACAGCAGCAGACAGATTGCTGAGTTTGGGGGAAGATGTTTC 2460  
 Oy 488 -----LeuGlnTrpHisAspProAlaThrArg 496  
 Db 2461 AGTGATTTCTGGGCAAAATACCTGCGCAGACCTGATCTCCGAGACCTCTGTGTGAT 2520  
 Oy 497 LysThrTyrProAsnGluProGluAlaLysLeuTyrLysArgIleAla 512  
 Db 2521 GACTCTCTC-----CAGCTGTGCAAGAAACAGCG 2550  
 RESULT 8  
 US-08-768-147B-1  
 : Sequence 1, Application US/08768147B  
 : Patent No. 5859222  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Leung, David W.  
 : APPLICANT: Tompkins, Christopher  
 : TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D  
 : NUMBER OF SEQUENCES: 14  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Cell Therapeutics, Inc.  
 : STREET: 200 Elliott Avenue West, Suite 400  
 : CITY: Seattle  
 : STATE: Washington  
 : COUNTRY: U.S.A.  
 : ZIP: 98119  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 3.5" disk, 1.44Mb, double side, high density  
 : COMPUTER: Hewlett Packard-IBM Compatible  
 : OPERATING SYSTEM: MS-DOS Version 6  
 : SOFTWARE: WORD for WINDOWS  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/768,147B  
 : FILING DATE: 16-Dec-1996  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: Provisional US60008768  
 : FILING DATE: 15-Dec-1995  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Faciszewski, Stephen  
 : REGISTRATION NUMBER: 36,131,  
 : REFERENCE/DOCKET NUMBER: 1802A  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (206)282-7100  
 : TELEFAX: (206)284-6206  
 : INFORMATION FOR SEQ ID NO: 1:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 3425  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single stranded  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: cdna to mRNA  
 : HYPOTHEICAL: no  
 : ANTI-SENSE: no  
 : FRAGMENT TYPE:  
 : ORIGINAL SOURCE:  
 : ORGANISM: homo sapien  
 : STRAIN:  
 : INDIVIDUAL ISOLATE:  
 : DEVELOPMENTAL STAGE:  
 : HAPLOTYPE:  
 : TISSUE TYPE:  
 : CELL LINE:  
 : ORGANELLE:

: IMMEDIATE SOURCE:  
 : LIBRARY:  
 : CLONE:  
 : POSITION IN GENOME:  
 : CHROMOSOME/SEGMENT:  
 : MAP POSITION:  
 : UNITS:  
 : FEATURE:  
 : NAME/KEY: hpcpld  
 : LOCATION:  
 : IDENTIFICATION METHOD:  
 : OTHER INFORMATION:  
 : PUBLICATION INFORMATION:  
 : AUTHORS:  
 : TITLE:  
 : JOURNAL:  
 : VOLUME:  
 : ISSUE:  
 : PAGES:  
 : DATE:  
 : DOCUMENT NUMBER:  
 : FILING DATE:  
 : PUBLICATION DATE:  
 : RELEVANT RESIDUES IN SEQ ID NO: 1:  
 : US-08-768-147B-1  
 :  
 : Alignment Scores:  
 : Pred. No.: 0.00102 Length: 3425  
 : Score: 121.50 Matches: 115  
 : Percent Similarity: 31.73% Conservative: 90  
 : Best Local Similarity: 17.80% Mismatches: 194  
 : Query Match: 4.48% Indels: 247  
 : DB: 2 Gaps: 30  
 :  
 : US-10-066-551-4 (1-525) x US-08-768-147B-1 (1-3425)  
 :  
 : Oy 30 LeuLeuCySserCysSerSer-----TrpLeuProLeuGluGlu 43  
 : Db 928 TTGATCTCAAGTCAGCAGCAGCTACCGCAGCAGCGTGCTGGCCCAAGATCATCTAG 987  
 : Oy 44 ArgThrGluSerArgHisAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63  
 : Db 988 CTGGCAGAGGC-----CCAGGCAGA-----GACTTCCTACAG 1020  
 : Oy 64 Ile---ArgHis-----ThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeu 79  
 : Db 1021 CTGCACCGGCATGACAGCTACGCCGCCCGCGCTGGACCTTGGCCCGGTGTTG 1080  
 : Oy 80 AsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSer 99  
 : Db 1081 AATGGGCGAGTGTACTTCTGCTGTCGTGCAGATGCCATCTTCAGCTCAGAGAGATT 1140  
 : Oy 100 LeuAspLeuGlnTyrTyrIle----- 106  
 : Db 1141 TTCATCAGACTGCTGTGTGAGTCTGAGTTTACCTGAGCGTCCGCCCATTCAGAT 1200  
 : Oy 107 ---TrpArgAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGlu 125  
 : Db 1201 GACTGAGACTGACATT-----ATGCTCAAGAGGAAGCGCGAG 1239  
 : Oy 126 ArgGlyValArgValArgLeuLeuAspAspAsnThrArgGlyLeuAspAspLeu 145  
 : Db 1240 GAGGCTGCTCGTGTCTATCTGTCTGCTTTAAGAA-----GTGCAATTG 1284  
 : Oy 146 LeuLeuAlaLeuAspSer-----HisProAsnIle 155  
 : Db 1285 GCCTTGGCATCAACAGTGCCTATACCAAGAACCGCTATGCTGTCACCAACATTA 1344  
 : Oy 156 GluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlnTyrLeuThr 175  
 : Db 1345 AAGTG---ATGCGTCAACCCAGACAGCAAGTACGTTGTG----- 1380  
 : Oy 176 AspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195

```

Db 1381 -----GCCCATCATGAGAGCTCCGTGTGGACCACTGGTA 1419
Oy 196 ThrIleuGlyGly----- 200
Db 1420 GCATTCCTGGGGGGAGCTGGAGCTTGCCTATGCGCGCTGGAGTACCTCCTACCGACTG 1479
Oy 201 ArgAniIleGlyAsp----- 205
Db 1480 ACTGACCTTGGAGACTCTCTGAATACAGCTCTCCAGCTCCACCGCGCCGAC 1539
Oy 206 -----GluTyrPheIysValGlyGluAsp----- 213
Db 1540 TCACCAAGCCACCCAGACCTCTGCACAAACCAATTTCTGCTGGCGCAAGAGCTACAGC 1599
Oy 213 ----- 213
Db 1600 AATCTTATACCAATGACTGGGTGAGCTGAGCCGCTTCGAAAGATTTCATTGACAGG 1659
Oy 214 -----ThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValVal 228
Db 1660 GAGAGACCCCTCGCATGCCATGCGGAGCTGGGGGTGTGCTGCATGCGCTACCGGCC 1719
Oy 229 GlyGluValSerHisAspPheAspArgTyrTrpAlaSerHisSerAlaHisAlaThr 248
Db 1720 CGGACCTTGGCCGGCACTTATCCAGCGCTG-----AACTTCACC 1761
Oy 249 ArgIleIleArgSerGlyAsnIleGlySgIleuAlaLeuGlyTyrAsnAspGlu 268
Db 1762 AAGACCAACCAAGGCC-----AACTACAGAGCTGCC 1791
Oy 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288
Db 1792 ACATACCCCTCGCTGCTCCAGCTACACAGCGGCAATCAGCTCCCTTC----- 1845
Oy 289 LysIleGlnThrGlyArgGlyIleAspTrpIleSerValGlnThrArgLeuIleSerAsp 308
Db 1846 ACATCTCCAGAGGGCACTGC-----ACCACGCTACAGCTC----- 1881
Oy 309 ProAlaLysGlyLeuAspArgAspArgArgLysProIleAlaGlyArgLeuGlnAsp 328
Db 1882 ---TTGGGATGATGAGACCGCTGTCA-----GCAGGAGCTCTGGAGAAC 1923
Oy 329 Ala-----LeuLysGlnProGluLysSerValTyrLeuVal 340
Db 1924 TCCATCTCAATGCCCTACCTGACACACATCCAGGGGAGCGACCTTCCTTACATTGAG 1983
Oy 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAsp 360
Db 1984 AATCAGTTCTTCACTAGCTGCTCAGATGGCGGACGGTTCTGAACAAAGGTGGCGATGAG 2043
Oy 361 GlyIleAsp-----ValThrValLeu 367
Db 2044 ATTGTGACAGAAATCTGAAGGCCCAACAAACAGGGGTGTGTTCACGAGTCTAGTGCTT 2103
Oy 368 ThrAsnSerLeuGlnAlaThrAsp-----Val 376
Db 2104 TTGCCCTTACTCTCTGCTTGCAGGGTGCATCTCAACGGCGGTGGCAATCCATCCAG 2163
Oy 377 AlaAlaValHisSerGlyTyr----- 383
Db 2164 GCCATTCTGCACTTACTTACAGAGCCCTGTGTGGGGAGTATTCAATCTTCATCCG 2223
Oy 384 -----ValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeu 397
Db 2224 CTTAAGACAGCCATGGGAGACAGATGGGAGATATATTTCATCTGCGGCTTCGTGACA 2283
Oy 398 TyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySer 417
Db 2284 CACGAGAGCTGGCGGCGACCCCGCTC-----TCG 2313
Oy 418 SerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGly 437

```

```

Db 2314 GAGCTCATCTACATCCACAGAGTGTCTATCGCAGATGACCGGAGATCATATTGAT 2373
Oy 438 SerPheAsnLeuAspProArgSer-----AlaArgLeuAsnThrGluMetGlyValVal 455
Db 2374 TCTGCAAAATCATCATATGACCGGAGCTTGTGTGGGAAAGCGGAGACATGAGCTGCTGTA 2433
Oy 456 IleGlu-----SerProLysIle-----AlaGluMetGluArgThr 468
Db 2434 ATCGAGCAGACAGAGAGAGCAATCCCTCATGAATGGCGAGAGTATCAGCGCGCAGG 2493
Oy 469 LeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArg 487
Db 2494 TTTCCTTGAAGTTCGCGGAGACCTGCTCAGTGTGATTTCTTGAGCAAAATCCCGCCA 2553
Oy 488 ---LeuGlnThrHisAspProAlaThrArgLysThrTyrProAsnGluProGluAlaLys 506
Db 2554 GACTTGATCTCCAGAGACCCCATCTGTGATGACTTCTTC-----CAG 2595
Oy 507 LeuTrpLysArgIleAla 512
Db 2596 TTGTGCAAGACATGGCT 2613

RESULT 9
US-09-107-149-1
: Sequence 1, Application US/09107149
: Patent No. 6274363
: GENERAL INFORMATION:
: APPLICANT: Leung, David W.
: TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
: FILE REFERENCE: 077319/0144
: CURRENT APPLICATION NUMBER: US/09/107,149
: EARLIER FILING DATE: 1998-06-30
: EARLIER FILING DATE: 1996-12-17
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 3425
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (64)..(2862)
US-09-107-149-1

Alignment Scores:
Pred. No.: 0.00102 3425
Score: 121.50 115
Percent Similarity: 31.73% Conservative: 90
Best Local Similarity: 17.80% Mismatches: 194
Query Match: 4.48% Indels: 247
DB: 4 Gaps: 30

US-10-066-551-4 (1-525) x US-09-107-149-1 (1-3425)
Oy 30 LeuLeuCySerCySerser-----TripleuProProLeuGluGlu 43
Db 928 TTGATTTCAAGTGCAGAGAGCTACCGGAGCGAGCGGTGTGGGCCCAAGATCATGAG 987
Oy 44 ArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeuGln 63
Db 988 CTGGCACAGGCG-----CCAGGCA-----GACTTCTTACAG 1020
Oy 64 Ile---ArgHis-----ThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeu 79
Db 1021 CTGCACCGGCACTGACAGTACGCCCCACCGCGCTGGAACTTGGCGCGGTGGTGTG 1080
Oy 80 AsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGlnHisSer 99
Db 1081 AATGGGAGAGGTACTTGTGCTGTGTGGAGATGCAATCTTCAGACTCAAGAGAGATT 1140
Oy 100 LeuAspLeuGlnTyrTyrIle----- 106

```

```

Db 1141 TTCATCAGACAGCTGTGGTTAGTCCGTGAGGTTTACCTGAAGCGGCCGCGCCATTGCAGAT 1200
      ::::::::::
Qy 107 ---TTTATGASnAspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleGlu 125
      |||||
Db 1201 GACTGGAGACTGAGCATT-----ATGCTCAGAGGAGGAGCGGAG 1239
      |||||
Qy 126 ArgGlyValArgValArgLeuLeuLeuAspAsnAsnThrArgGlyLeuAspAsnLeu 145
      |||||
Db 1240 GAGGTTCCCTGTCTCTATCTCTCTGTCTTTAAGAA-----GTGGAATG 1284
      |||||
Qy 146 LeuLeuAlaLeuAspSer-----HisProAsnIle 155
      |||
Db 1285 GCCTTGGGATCAGACAGGCTATAGCAGAGGCGCTGATGCTGCTGCACCCCAACATA 1344
      |||
Qy 156 GluValArgLeuPheAsnProPheValLeuArgLysTyrPheAlaLeuGlyTyrLeuThr 175
      |||
Db 1345 AAGGTG---ATGCTCCACCCAGACCAAGTGCCTTTGCG----- 1380
      |||
Qy 176 AspheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195
      |||
Db 1381 -----GCCATCATGAGAAAGCTCCGTGTGTGTGAGACCAAGTGTGTA 1419
      |||
Qy 196 ThrIleLeuGlyGly----- 200
      |||
Db 1420 GCATTCTGGGGGAGCTGAGCTTGCCTATGCGCGTGGATGAGACTGCATCCGACTG 1479
      |||
Qy 201 ArgAsnIleGlyAsp----- 205
      |||
Db 1480 ACTGACCTTGAGACTCCTCTGTAATCAGCTCCGCCAGCTCCACCCCGCGCCAGAC 1539
      |||
Qy 206 -----GluTyrPheLysValGlyLysP----- 213
      |||
Db 1540 TCACAGCAGCAGCCAGACCTCTCTCACAACCAATTTCTGTGCTGGCAGAGACTACAGC 1599
      |||
Qy 213 ----- 213
      |||
Db 1600 AATCTTATCACAATGACTGGTGCAGCTGAGCGGCTTTCGAAGATTTCATTGACAGG 1659
      |||
Qy 214 -----ThValPheAlaAspLeuAspIleLeuAlaIleGlySerValVal 228
      |||
Db 1660 GAGACGACCCCTCGATGCCATCCAGCGGAGCGTGGGTGCTGCATCGCTTACCGGCC 1719
      |||
Qy 229 GlyLeuValSerHisAspPheAspArgTyrTrrPalaSerHisSerAlaHisAsnAlaThr 248
      |||
Db 1720 CGGAGCTTGGCCGCGCACTTATCCAGCGCTG-----AACTTCACC 1761
      |||
Qy 249 ArgIleIleArgSerGlyAsnIleGlyLysGlyLeuAlaLeuGlyTyrAsnAspGlu 268
      |||
Db 1762 AAGACACCAAGGCC-----AAGTACACAGACTCCC 1791
      |||
Qy 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288
      |||
Db 1792 ACATAACCCCTACCTCCCAAGCTACACAGAGCGGCAATCAGCTCCCTTC----- 1845
      |||
Qy 289 LysIleGlnThrGlyArgIleAspTrrPleInSerValGlnThrArgLeuIleSerAspAsp 308
      |||
Db 1846 ACACCTCCAGAGGCGCAGTGC-----ACACCTGACAGGTC----- 1881
      |||
Qy 309 ProAlaIleGlyLeuAspArgAspArgArgLysProPheIleAlaGlyArgLeuGlnAsp 328
      |||
Db 1882 ---TTGCGATCAGTGAGCGCTGTGCA-----GAGGAGACCTGTGAGAAC 1923
      |||
Qy 329 Ala-----LeuLysGlnProGluLysSerValTyrLeuVal 340
      |||
Db 1924 TCCATCTCAATGCCCTACCTGCACACCACTCAGGGGAGCCAGCACTTCTTACATTGAG 1983
      |||
Qy 341 SerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAsp 360
      |||
Db 1984 AATCAGTTCTTATTAGTCTGCTCAGATGCGGAGCGTTTGCACAAAGGCGGCATGAG 2043
      |||
Qy 361 GlyIleAsp-----ValThrValLeu 367
      |||

```

```

Db 2044 ATTGTGACAGAAATCTGTGAAGCCCAACACAGGGGTGTGTATACGAGTACGTCTT 2103
      |||
Qy 368 ThrAsnSerLeuGlnAlaIleThrAsp-----Val 376
      |||
Db 2104 TTGCCCTTACTCCCTGGCTTGCAGGGTGACATCTCCACGGGCGGTGCAGATCCATCCAG 2163
      |||
Qy 377 AlaIleValHisSerGlyTyr----- 383
      |||
Db 2164 GCCATTCTGCACCTTACTTACAGACCCCTGTGCTGGGAGATATTCATCTGCATCCG 2223
      |||
Qy 384 -----ValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeu 397
      |||
Db 2224 CTTAAGCAGCCATGGGAGCAGCATGGCGGACTATATTCCATCTGCGGCTTCGTACA 2283
      |||
Qy 398 TyrGluLeuGlnProAsnHisAlaValProAlaIleThrLysAspLysGlyLeuThrGlySer 417
      |||
Db 2284 CACGAGAGCTGGGGGCGACCCCGTC-----TCG 2313
      |||
Qy 418 SerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGly 437
      |||
Db 2314 GAGCTCATCTACATTCACAGCAGAGGTGCTCATCGCAGATGACCGGACGTCATCATTCAT 2373
      |||
Qy 438 SerPheAsnLeuAspProArgSer-----AlaArgLeuAsnThrGluMetGlyValVal 455
      |||
Db 2374 TCTGCAAAATCAATATGACCGGAGCTTGTGCTGGGGAAGCGGACAGTGAAGCTGGCGTCTGA 2433
      |||
Qy 456 IleGlu-----SerProLysIle-----AlaGluGlnMetGluArgThr 468
      |||
Db 2434 ATCGAGGACACAGACGAGAACCATCTCTCATGAATGGGCGCATATACAGGCGGCGCAGG 2493
      |||
Qy 469 LeuAlaSerThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArg--- 487
      |||
Db 2494 TTGGCTTGATCTGCGGAGAACACATCGCTTCACTGATGATCTTGGAGCAATACCGGCA 2553
      |||
Qy 488 ---LeuGlnThrPheAspProAlaIleThrArgLysThrTyrProAsnGluProGluAlaLys 506
      |||
Db 2554 GACTTGGATCTCCGAGACCCCATCTGTGATGACTTCTTC-----CAG 2595
      |||
Qy 507 LeuTrpLysArgIleAla 512
      |||
Db 2596 TTGTGGCAAGACATGGCT 2613
      |||

RESULT 10
US-08-961-527-5/c
: Sequence 5, Application US/08961527
: Patent No. 6420135
: GENERAL INFORMATION:
: APPLICANT: Charles Kunsch
: TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
: NUMBER OF SEQUENCES: 391
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
: COMPUTER: HP Vectra 486/33
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,527
: FILING DATE:
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brookes, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P1

```

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (301) 309-8504  
 TELEFAX: (301) 309-8512  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 13926 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 US-08-961-527-5

## Alignment Scores:

Score:	113.50	Length:	13926
Percent Similarity:	35.08%	Matches:	119
Best Local Similarity:	23.06%	Conservative:	62
Query Match:	4.18%	Mismatches:	191
		Indels:	145
		Gaps:	27

US-10-066-551-4 (1-525) x US-08-961-527-5 (1-13926)

```

QY 41 LeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAspAsn 60
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4464 CTCAGCTCTCAACGAGTATCAGACATCTAATGATAGGTAGTACATAGAGT----- 4411

QY 61 IleuGlnIleuArgHisThrProHisThrAsnGlyLeuSerAspIleuTyrrLeuLeuAsn 80
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4410 -----AATGCATCAAGATTTATATTA----- 4387

QY 81 AspProHisGluAlaPheAla-----AlaArgAlaIleu 92
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4386 -----GACAAAGCTTCACAGATGTTAAAGAAACTTAACAAACCTTAATACCAAAATTA 4333

QY 93 IleGluSerAlaGluHisSerLeuAsp-----LeuGlnTyr 104
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4332 GTTCAAAACGACAAACATCAATTAATGATTCCTGCTGCTCAATGATTCGAT 4273

QY 105 TyrrIleuArgHisAsnIleSerGlyArgLeuLeuPheAsnLeuValTyrrLeuAlaAla 124
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4272 AAGTTCAGAAAC-----AAGCAGCTTATTACTAGGTTAACTTACCTCAATATCGT 4219

QY 125 GluArgGlyValArgVal-----ArgLeuLeuLeuAspAspAsn 138
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4218 TACTATGAGATTAAATTGGTATGATTAATTAACAATTAATGCTATTAACAACGAT 4159

QY 139 ThrArgGly-----LeuAspAspLeuLeuAlaLeuAspSerHisPro 153
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4158 TTCTATGTCGAAAAAGTTAGCGTATTTAGACAGATTATTAATCGGTTCTTAAGAGAAC 4099

QY 154 AsnIleGlu-----ValArgLeuPheAsnProPhe-----ValLeuArgGlySTPRArgAla 170
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4098 AACATTAAAGCTTCACCTACATTCGACGCAATTCGGTCAAGTATGGCTAATATATCTAAA 4039

QY 171 LeuGlyTyrLeuThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThr 190
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 4038 TCAGGTATTTAGATGATTTTAAATTATATAGACAATG-----TTTACA 3991

QY 191 AlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlyTyrPheLysVal 210
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3990 AATATAGAC-----AATATGACGATGTTGGTATTTAGATGCT 3955

QY 211 GlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrArgLysValValGlyGlu 230
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3954 ACAAGAACCATGCTAC-----ATCGACAGACCGCGCTTGAAGTCGAGAA 3907

QY 231 ValSerHisAsp-----PheAspArgTyrTrpAlaSerHisSerAlaHisAsn 246
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3906 ATTAATAATCTTAACATCTGCTGATTCGATTAATTAACAAGAACTCCCTTAGAAATACT 3847

QY 247 AlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeu-----Gly 264
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3846 ATACTCCCATCTAGC-----AATATGATTAAGACACATCTTTATTTAATTCAAT 3796
  
```

```

QY 265 TyraAspGluThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGln--- 283
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3795 TATATCAATTCGCTTGGTAGTCAGACGATAGTAAATAATCATTAAGATAT 3736

QY 284 SerProLeuTyrrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArg 303
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3735 AAAGATATCGTTAAACAAGCTGCAGATGTTATAGAAACATTAATGATTTCTGTATCTG 3676

QY 304 LeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysProProIleAla 323
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3675 CTAGCGTCTATACGTTAAACAAGCTACTAAGATCTGTATTCCTATTGGAA 3616

QY 324 Gly-----Arg-LeuGlnAspAlaLeuLysGln-----ProGluLysSe 336
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3615 GGTATATACCTCTGCTGATGATGTTGAATAAATATGCGCTAATACCGCAAGATA 3556

QY 336 ValTyrrLeuValSerProtyrPheValPro-----ThylSerGlyThr 351
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3555 TATATCTCTTATAGAA--TTCTTTGGCTCTATGATAGTATTAATTAATATGAAAC 3497

QY 351 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerIle 371
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3496 AGGAGCTTATGCTGCTATATATATCTAATCTGATGATAT----- 3457

QY 371 uGlnAlaThrAspValAlaAlaValHisSerGlyTyrrValLysTyrrArgLysProLeuLe 391
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3456 -----AGAACTGATGTAAATATGTTCAATTAGAATGCTGGT----- 3418

QY 391 uLysAlaGlyIleLysLeuTyrrGluGlnProAsnHisAlaValProAlaThrLysAs 411
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3417 -GAATACGATTTTACGTTTACACACATGCAACACACAC----- 3379

QY 411 PLYSGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspG 431
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3378 ----- 3371

QY 431 YLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrG 451
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3370 CCGTGCATTACTTACTAGTGGCTTT-----GGACACCGTCAAGCTACT-- 3328

QY 451 uMetGlyValValIleGluSerProLysIleAlaGluGlnMetGluArgThrLeuAla-- 470
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3327 -----GATCTGAGACATATGCTCAAGGATATCTCAAAACTCTCTTAC 3284

QY 471 -----AspThrSerProGluTyrrAlaTyrrArgValThrLeuAspArgHisAs 486
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3283 TGGTAGTGAATTGATGACAGTTGGTTCTTATGATTAATATGCTATTAAACGCAAAA 3224

QY 486 n-----ArgLeuGlnTriHis-----AspProAlaThrArgLysThr 498
    |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
Db 3223 TGATGGAAATCAGTGTATATTTACAGATCCAAAACCTTAATAACA 3178

RESULT 11
US-08-968-752B-3
Sequence 3, Application US/08968752B
Patent No. 6043073
GENERAL INFORMATION:
APPLICANT: Frohman, Michael A.
TITLE OF INVENTION: No. 6043073el Phospholipase D Polypeptide and
TITLE OF INVENTION: DNA Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: ONYX Pharmaceuticals, Inc.
STREET: 3031 Research Drive
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94806
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
  
```







```

Db 2299 ATCTGAAGCTCAGAGGAAACAGAAATACCGGATATGCTGATACCA---CTT 2355
Qy 345 ValPro-----ThlySerglyThrAspAlaLeuAlaLysLeuVal 358
Db 2356 CTGCCAGGCTTCGAGAGACATTTCAACCGCGAGAGAAATGCTCTCAGGCATCATG 2415
Qy 359 GlnAspGlyLeuAspValThrValLeu---ThrasnSerLeuGlnAlaThrAspValAla 377
Db 2416 CACTTCAACTACACACACCATCTGCAGAGAGAGAAATTCATCTTCCTGGACAGTTAAAGCA 2475
Qy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyTleLysLeu 397
Db 2476 GAGCTTGCTATCAGTGGATAAATTAC-----ATATCATCTCTGGCTTGA--- 2523
Qy 398 TyrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySer 417
Db 2524 -----ACACATGCA-----GACCTCGAAGGAAAC 2547
Qy 418 SerValThrSerLeu-----HisAlaLysThrPheIleValAspGlyLysArgTle 434
Db 2548 CTAGTAACTGAGCTTATCTATGCTCCACAGCAAGTTGTTAATGCTGATGATACACTGTT 2607
Qy 435 PheIleGlySerPheAsnLeuAspProArgSer-----AlaArgLeuAsnThrGluMet 452
Db 2608 ATATTGGCTCTGCCAACAATAATGACCGCAGCATGCTGGGAAAGCGTGACAGTGAATG 2667
Qy 453 GlyValValIleGlySer-----ProLysIleAlaGluGlnMetGluArgThr 468
Db 2668 GCTCTCATTTGCAAGATACAGACAGCTGCTTCAGTAATGAT--- 2712
Qy 469 LeuAlaAspThrSerProGlyTyr----- 476
Db 2713 -----GGAAAGAGTACCAACACTGCGCGGTTGCCCGAGACTTCGGCTACAG 2760
Qy 477 AlaTyrArgValThrLeu-----AspArgHisAsnArgLeuGlnThrHisAsp 492
Db 2761 TGCTTAGGTTGCTTCCTGCTGATCTGATGACCAAGTGAGAGACATTCAG-----GAT 2814
Qy 493 ProAlaThrArgLysThrTyrProAsnGluProGluAlaLysLeuThrLysArgTleAla 512
Db 2815 CCAGTGAAGTGAACAAATCTCTC-----AAGGAGGCTGTGGTTTCAACACACA 2859
Qy 513 Ala-----LysIleLeuSerLeuLeuPro 520
Db 2860 GCTGCAATGCTACAAATTATGACAGAGTTTCCGCTGCTTCCC 2904

RESULT 13
US-08-073-384C-21
: Sequence 21, Application US/08073384C
: Patent No. 5541311
: GENERAL INFORMATION:
: APPLICANT: Dahlberg, James E.
: APPLICANT: Lyamlchev, Victor I.
: APPLICANT: Brow, Mary Ann D.
: TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
: TITLE OF INVENTION: POLYMERASE
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/073.384C
: FILING DATE: 04-JUN-1993

```

```

: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/986,330
: FILING DATE: 07-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Carroll, Peter G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-00613
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/705-8410
: TELEFAX: 415/397-8338
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2502 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-073-384C-21

Alignment Scores:
Pred. No.: 0.041 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
DB: 1 Gaps: 30

US-10-066-551-4 (1-525) x US-08-073-384C-21 (1-2502)
Qy 118 AsnLeuValTyrLeuAlaAlaGluArgLysValArgVal----- 130
Db 961 GATCTTCTGGCCCTGCGCCGCCAGAGGGGGGGCGGTCCACCGGGCCCCGAGCTTAT 1020
Qy 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAAGCCCTCAGGAGCCTGAGAGGCGGGGGCTTCTCGCAAGACCTGAGCGTTCTG 1080
Qy 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGGAAGGCTTGGCTCCCGCCCGCGAGCAGCCCATGCTC----- 1128
Qy 164 ValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTGCGTACCTCTCGAGCCCTTCACACACACCCCGAG 1167
Qy 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGGCCCGCGCTACGCGCGGAGTGGAGGAGCGGGGAGCGCGCCCTT 1227
Qy 198 -----LeuGlyArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGGCTTTCGCCACACTGTGGGGAGGCTTGAAGGGAGAGAGGCTCCTTGG 1287
Qy 207 ---TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACCGGAGGTGGAGAGGCGCCCTTCCGCTGTCCTGCCACATGAG-----GCC 1341
Qy 224 ThrGly-----SerValAlaGlyGlu 230
Db 1342 ACGGGGTGCGCGCTGAGCTGACCTATCTCAGGGCTTGTCCCTGAGAGTGGCCGGGGAG 1401
Qy 231 ValSer-----HisAspPheAspArgTyrTrp 239
Db 1402 ATCGCCCGCTCGAGCGCCAGGCTTCCCGCTGGCCGCGCCACCCCTTCAC----- 1452
Qy 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259
Db 1453 ---CTCACTCCCGGAGACAGCTGMAAGGTCCTTTCAG-----GAGTAGGG 1500
Qy 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCCCGCATGCGCAAGAGAGAGAACCGGCGCTCCACAGCGCCGCGCTCTG 1560

```

Oy 274 -----LeuAaArgTyrTrpGluThrVal 281  
 Db 1561 GAGGCCCTCCGAGGAGGCCACCACCATCGTGTGAGAAAGATCTCTGCAGTACCGGGAGCTCAC 16220  
 Oy 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294  
 Db 1621 AAGCTGAGAGACACCTACATGTAGACCCCTTCGCGAGCATCTTCCACACCCAGAGACGGGCGCG 1680  
 Oy 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAsp 309  
 Db 1681 CTCACACCCCTTCCACACACGCGCCAGCGCCAGGGCAGGGCTAGTGTGCTCCATGCC 1740  
 Oy 310 AlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyArgLeuIlnsPala 329  
 Db 1741 -----AACCTCCAGAACATCCCGCTCCGACCCCGCTTGGAGAGAGATCCGCGGGCC 1790  
 Oy 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343  
 Db 1795 TTCATCGCCGAGGAGGGGTGGCTATTGTGTGGCTGAGACTAGCCAGATAGACCTCAG 1850  
 Oy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspLysIleAsp 363  
 Db 1855 GTGCTGGCCCACTCTCCGGGAGAGAACCTATCGGATCTTCCAGGAGGGGGGAG 1914  
 Oy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspAla 377  
 Db 1915 ATCCACAGGAGACCGCAGCTGATGTGGGGTCCCGGAGAGCCCTGGAC----- 1960  
 Oy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393  
 Db 1969 -----CCCTGATGGCGCGCGCCGACCAAGC 1999  
 Oy 394 -----GlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406  
 Db 1996 ATCAACTTCGGGGTCTCTACGCGATGTGCGCCACGCCCTCTCCAGAGCTACCATC 2055  
 Oy 407 Pro----- 407  
 Db 2056 CCTTACGAGGAGGCCCAAGGCTTCATTGACGCGTACTTTCACAGCTTCCCAAGTGTGGG 2115  
 Oy 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423  
 Db 2116 GCCTGGATTGAGAAAGACCTCGAGAGAGGAGGCGGGGGTACGTGAGACCTCTTC 2179  
 Oy 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443  
 Db 2176 -----GGCGCGCGCGCTACGTCACAGCTTACAGGCCCGGGTG 2214  
 Oy 444 ArgSerAlaArgLeuAsnThrGlnMetGlyValIleGluSerPro-----Lys 460  
 Db 2215 AAGAGCGCGGAGGCGCGCGAGAGCGC--ATGGCTTCACATGCGCGTCCGGGGCAC 2271  
 Oy 461 IleAlaGlnMetGluAlaGlyThrIleuAlaAspThrSerPro-----GluTyrAla 477  
 Db 2272 GCGCGCGACTCAAGAGCTGCGTATGTGTAAGACTTCCACAGCTGTGAGGAATAATGGG 2331  
 Oy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpPheIleAspProAlaThrArgLys 497  
 Db 2332 GCCAGGATGCTCTTTCACAGCTCCACGAGAGACTGTCTTGAGGCCCCCAAGAGAGGGCG 2391  
 Oy 498 ThrTyrProAsnGluProGluAlaLysLeuThrLysArgIleAlaAlaLysIleLeuSer 517  
 Db 2392 -----GAGGCGCGGGCGCGCGCTGCCAAGAGAGTATGAGAGGGGTATATCC 2439  
 Oy 518 Leu---LeuProIleGlu 522  
 Db 2440 CTGGCGCGTGGCGCGTGGAG 2457

RESULT 14  
 US-08-254-359A-21  
 ; Sequence 21 Application US/08254359A  
 ; Patent No. 5614402

```

1 GENERAL INFORMATION:
2 APPLICANT: DAHLBERG, JAMES E.
3 APPLICANT: LYAMICHEV, VICTOR I.
4 APPLICANT: BROW, MARY ANN D.
5 TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
6 TITLE OF INVENTION: DNA POLYMERASE
7 NUMBER OF SEQUENCES: 40
8 CORRESPONDENCE ADDRESS:
9 ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
10 STREET: 220 MONTGOMERY STREET, SUITE 2200
11 CITY: SAN FRANCISCO
12 STATE: CALIFORNIA
13 COUNTRY: UNITED STATES OF AMERICA
14 ZIP: 94104
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patent Release #1.0, Version #1.25
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/254,359A
23 FILING DATE:
24 CLASSIFICATION: 435
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 08/073,384
27 FILING DATE: 06-JUN-1993
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 07/986,330
30 FILING DATE: 07-DEC-1992
31 ATTORNEY/AGENT INFORMATION:
32 NAME: CARROLL, PETER G.
33 REGISTRATION NUMBER: 32,837
34 REFERENCE/DOCKET NUMBER: F08S-01000
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (415) 705-84101
37 TELEFAX: (415) 397-8338
38 INFORMATION FOR SEQ ID NO: 21:
39 SEQUENCE CHARACTERISTICS:
40 LENGTH: 2502 base pairs
41 TYPE: nucleic acid
42 STRANDEDNESS: double
43 TOPOLOGY: linear
44 MOLECULE TYPE: DNA (genomic)
45
46 US-08-254-359A-21
47
48 Alignment Scores:
49 Pred. No.: 0.041 Length: 2502
50 Score: 106.00 Matches: 119
51 Percent Similarity: 33.70% Conservative: 65
52 Best local Similarity: 21.79% Mismatches: 174
53 Query Match: 3.91% Indels: 188
54 DB: 1 Gaps: 30
55
56 US-10-066-551-4 (1-525) x US-08-254-359A-21 (1-2502)
57
58 QY 118 AsnLeuValTyrLeuAlaIaGluArgGlyValArgVal----- 130
59 ::::::::::: ||||||| ||||| |||||
60 Db 961 GATCTTGTGCGCCGTGGCGCCGCGCAGGAGGGGGCGGTCCACCGGGCCCCCAGCCTTAT 1020
61 ArgLeuLeuAspAspAsnAsnThrArgGlyLeu-----AspAspLeuLeu 147
62 :::: ||| ||| ||||||| ||| ::::|
63 Db 1021 AAGCCCTCAGGACGACGAGGAGGCGCGGCTCTCGCAAGACCTGACGCTTCG 1080
64 AlaLeuAsp-----HisProAsnIleGluValArgLeuPheAsnProHe 163
65 ||||| ||| ||| :::: |||
66 Db 1081 GCCCTTGGAGGAGAGCGCTTGCCCGCCGCGCGACGACCCATGCTC----- 1128
67 ValLeuArgTyrPArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
68 ||| |||||| |||
69 Db 1129 -----CTCGGCTTACTCTCTGACCTTCACACGACGCCCGAG 1160
70 ||| ::::| |||
71 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
72 ||| :::: ||| ||||| :::

```

```

Db 1168 GGGGTGCCCCGAGCTACGCGGGGAGTGAAGAGAGGCGGGGAGCGGCGCCCTT 1227
Qy 198 -----LeuGlyIyArgAnIleGlyAspGlu-----Lys 206
Db 1228 TCCGAGAGGCTTCCGCAACCTGTGGGGAGAGCTTGAAGGGAGAGAGCTCTTGG 1287
Qy 207 ---TyrPheIysValGlyIle-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACCGGGAGGTGAGAGAGGCGCCCTTCCGCTGTCCGCTGCGCCACATGAG-----GCC 1341
Qy 224 ThrGly-----SerValIleGlyIle 230
Db 1342 ACGGGGTGGCGCTGAGAGCTGAGCTATCTACAGGCTTGTCCCTGAGAGTGGCGGGAG 1401
Qy 231 ValSer-----HisAspPheAspArgIyTrp 239
Db 1402 ATCGCCGCGCTGAGGCGGAGGTCTTCCGCTGCGCGGACCGCTTCAC----- 1452
Qy 240 AlaSerHisSerAlaHisAlaThrArgIleIleArgSerGlyAsnIleGlyIysGly 259
Db 1453 ---CTCAACTCCCGGAGACCTGGAAGGGCTCTTTCAC-----GAGCTAAGG 1500
Qy 260 LeuGlnAlaLeuGlyIyTrpAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCGCGCATGCGAGAGAGCGAGAACCGGCAAGCGCTCCACGCGCGCGCTCTG 1560
Qy 274 -----LeuLeuArgIyTrpArgGluThrVal 281
Db 1561 GAGCGCTCCGCGAGGCGCCACCCCTGCTGAGAGATCTGCACTACCGGGAGCTCAC 1620
Qy 282 GluGln-----SerProLeuIyTrpGlnIysIle-----GlnThrGlyArg 294
Db 1621 AAGCGAGAGAGACCTACATATGACCCCTTGGCGGACCTCATCCACCCAGAGCGGCGCG 1680
Qy 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309
Db 1681 CTCACACCGCGCTTCACCAAGCGAGCGCCACGCGGAGCTAAGTCCATCC 1740
Qy 310 AlaIysGlyLeuAspArgAspArgIyTrpProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCCAGACATCCCGCTCCGACCGCTTGGGAGAGATCGCGGCGC 1794
Qy 330 LeuIysGlnProGluIySerValIyTrpLeuValSerProIyTrp----- 343
Db 1795 TTCATCGCGGAGAGGGGTGCTATGTTGGCGCTGAGCTATAGCCAGATAGAGCTCAG 1854
Qy 344 PheValIProThrIySerGlyThrAspAlaLeuAlaIysLeuValGlnAspIyIleAsp 363
Db 1855 GTGCTGGCCCACTCTCCGCGAGAGAACCTGATCCGGGTCTCCAGAGAGGGCGGAGC 1914
Qy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACAGGAGAGCGGACGCTGATGTTCCGCGCTCCCGGAGGCGCTGAGC----- 1968
Qy 378 AlaValHisSerGlyIyTrpValIyTrpArgIyTrpProLeuLeuIysAla----- 393
Db 1969 -----CCCTGATCGCGCGCGCGCGCGCAAGACC 1995
Qy 394 -----GlyIleIysLeuIyTrpGlnLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTCGGGGTCTCTACGCGATGTCCGCCACCGCTCTCCAGAGAGCTACCCAGC 2055
Qy 407 Pro----- 407
Db 2056 CTTTACGAGAGAGCGGCTTCATGAGCGCTTCTACAGCTTCCCAAGGTGGG 2115
Qy 408 -----AlaThrIysAspIySerGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCCTGATTTAGACAGACCTGTGAGAGAGGCGGAGGAGGCTGAGTGGAGACCTCTTC 2175
Qy 424 AlaIyThrPheIleValAspGlyIySerArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GGCGCGCGCGCTACGTCCAGACCTAGAGCGCGGCTG 2214

```

```

Qy 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerPro-----Lys 460
Db 2215 AAGAGCGTGGCGGAGCGCGCGAGCGC---ATGCGCTTCACATGCGCCGTCCGGGCGAC 2271
Qy 461 IleAlaGluMetGluArgThrLeuAlaAspThrSerPro-----GluTrpAla 477
Db 2272 GCCCGGACCTCATGACGTGGCTATGGTGAAGCTTTCGCCAGGCTGGAGCAATGGGG 2331
Qy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnIyTrpHisAspProAlaThrArgIys 497
Db 2332 GCCAGATGCTCTTCAGGTCCACGAGCGAGCTGCTCGAGGCGCCCAAGAGAGGGCG 2391
Qy 498 ThrTrpProAsnGluProGluAlaIySerTrpIySerArgIleAlaIalIyIleLeuSer 517
Db 2392 -----GAGCGCGTGGCGCGCGCTGGCCAGAGAGGTATGAGAGGGGTGTATCCC 2439
Qy 518 Leu---LeuProIleGlu 522
Db 2440 CTGCGCGTGGCGCGCGCTGAG 2457

RESULT 15
US-08-483-043-21
: Sequence 21, Application US/08483043
: Patent No. 5691142
: GENERAL INFORMATION:
: APPLICANT: Dahlberg, James E.
: APPLICANT: Lyamichev, Victor I.
: APPLICANT: Brow, Mary Ann D.
: TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESS: HAYESSTOCK, MEDLEN & CARROLL
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,043
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/073,384
: FILING DATE: 04-JUN-1993
: APPLICATION NUMBER: US 07/986,330
: FILING DATE: 07-DEC-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Carroll, Peter G.
: REGISTRATION NUMBER: 32,837
: REFERENCE/DOCKET NUMBER: FORS-00613
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415/705-8410
: TELEFAX: 415/397-8338
: INFORMATION FOR SEQ ID NO.: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2502 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-483-043-21

Alignment Scores:
Pred. No.: 0.041 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65

```

Best Local Similarity:	21.79%	Mismatches:	174
Query Match:	3.91%	Indels:	188
DB:	1	Gaps:	30

QY	118	AsmLeuVal1TyrLeuAla1IaGluATrGlyValATrVal	-----	AspAspLeuLeu	147
Dp	961	GATCTTGTGGCCCTGGCCCGCCGACAGGGGGCGGGGTCCACCGGGCCCCGACGCTTAT	-----		102
QY	131	ArgLeuLeuLeuAspAspAsnThrATrGlyLeu	-----	AspAspLeuLeu	147
Dp	1021	AAACCCCTTCAGGGACCTGAAGAGCGCGGGGGCTTCTGCCAATAACCTGAGCTTGG	-----		1080
QY	148	AlaLeuAspSer	-----	HisProAsn1IleGluValATrGlyLeuAsnProPhe	163
Dp	1081	GGCCTGAGGGAAGGCTTGGCTTCGCCCGCGGAGGAGGCCATGGTC	-----		1128
QY	164	ValLeuATrGlyStrPArGluAlaLeuGlyTyrLeuThrAsp	-----	PheProArg	179
Dp	1129	-----CTCGGCTACCTCTCGACCTTGCACACACACCCCGAG	-----		1167
QY	180	LeuAsnATrGlyMetHisAsnLysSerPheThr	-----	AlaAspAsnATrGluThrIle	197
Dp	1168	GGGGTGGCCCGCGCTACGGCGGGGAGTGGACCGAGAGCGGGAGCGGGCCGCTT	-----		1222
QY	198	-----LeuGlyGlyATrGlyAsn1IleGlyAspGlu	-----		206
Dp	1228	TCCGAGAGGCTTTCGCCAACCTTGGGGAGGCTTGAAGGGGAGAGAGAGCTCTTGG	-----		1287
QY	207	---TyrPheLysValGlyGlu	-----	AspThrValPheAlaAspLeuAsp1IleLeuAla	223
Dp	1288	CTTTACCGGAGGTGGAGAGAGCCCTTTCGGCTGTCTGGCCACATGAG	-----	-GCC	134
QY	224	ThrGly	-----	SerValValGlyGlu	230
Dp	1342	ACGGGGGTGGCCTGGAGCTGGCTATCTAGGGCTTGTCCCTGGAGGTGGCGGGAG	-----		1401
QY	231	ValSer	-----	HisAspPheAspATrGlyTyr	239
Dp	1402	ATGCGCCGCTCGAGGCGGAGGTCTTCGCGCTGGCGCGGACCCCTTCAC	-----		1452
QY	240	AlaSerHisSer1IaHisAsnAlaThrATrGly1IleATrGlySerGlyAsn1IleGlyLysGly	-----		259
Dp	1453	---CTCACTCCCGGAGCACAGCTGGAAAGGGTCTCTTGAC	-----	-GAGCTAAGG	1500
QY	260	LeuGluAlaLeuGlyTyrAsnAspGluThrSerATrGlyHisAla	-----		273
Dp	1501	CTTCCGGCATCGGCAGACGGAGAGACGGGAGCGCTCCACACAGCGCGCGCTCG	-----		1560
QY	274	-----LeuLeuATrGlyATrGlyGluThrVal	-----		281
Dp	1561	GAGGCCCTCGCGAGGCCACCCCATCTGTGAGAAATCTCTCAATACCGGGAGCTCAC	-----		1620
QY	282	GluGlu	-----	SerProLeuTyrGluLysIle	294
Dp	1621	AACTGAAAGACACTACATTCAGCCCTTGGCGGACATCATCCACCCAGACGGCGCC	-----		1680
QY	295	IleAspTrpGlnSerValGlnThr	-----	ArgLeu1IleSerAspAspPro	309
Dp	1681	CTCCACACCCGCTTCAACCAAGCGGCACGGCCACGGGAGGCTAATGATCTCCGATCC	-----		1748
QY	310	AlaLysGlyLeuAspATrGlyAspATrGlySerProPhe1IleAlaGlyATrGluGlnAspAla	-----		329
Dp	1741	-----AACTCCAGAAATCCCGCTGCCACCCCGCTTGGCGAAGATTCGCCGGGCC	-----		1794
QY	330	LeuLysGlnProGluLysSerValTyrLeuValSerProTyr	-----		343
Dp	1795	TTCATGCGCCAGAGGGGTGGCTATGTTGGTGGCCCTGACATTAAGCAAGATAGCTCAG	-----		1854
QY	344	PheValProThrLysSerGlyTyrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp	-----		363
Dp	1855	GTGCTGGCCACTCTTCGCGCGACGAACCTGTATCCGGGTCTTCCAGAGGGGGGAG	-----		1914

QY 364 ValThrValLeuThrIAsnSer-----LeuGlnAlaThrSpAla 377  
 Db 1915 ATCCACACGAGAGACCCTGACTGATGTTCCGGGTCCCGGGAGAGCCCTGAC----- 1968  
 QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393  
 Db 1969 -----CCCTGATCGCGCGGCGCCAAAGC 1995  
 QY 394 -----GlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406  
 Db 1996 ATCACTTCGGGGTCTCTACAGCGCATGTGGCCACCGCTCTCCAGAGACTACCAATC 2055  
 QY 407 Pro----- 407  
 Db 2056 CCTTACAGAGAGGCCACGAGCCTTCATTGAGCGCTACTTCAGAGCTTCCCAAGTGGCG 2115  
 QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423  
 Db 2116 GCGTGTATGTAGAAGACCCCTGAGGAGAGGCGAGAGCGGGGTACGTAGACACCTCTTC 2175  
 QY 424 AlaLysThrThrIleValAspGlyLysArgGlyIlePheIleGlySerPheAsnLeuAspPro 443  
 Db 2176 -----GCCCGCCCGCGCTACGTCCAGACCTTAAAGGCCCGGCTG 2214  
 QY 444 ArgSerAlaArgLeuAsnThrGlnMetGlyValIleGluSerPro-----Lys 460  
 Db 2215 AAGAGCGTGGGGAGGGGCGCCAGAGCC---ATGCGCTTACACATGCTCCGTCGGGCGAC 2271  
 QY 461 IleAlaGlnMetGlnArgThrLeuAlaAspPheSerPro-----GluTyrAla 477  
 Db 2272 GCCCGCCACCTCATGAAGCGGTGATGATGATGAAGACCTTCTCCAGCGCTGAGGAAATGGCG 2331  
 QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497  
 Db 2332 GCCAGGATGCTCTTCAGGTCACAGACAGCTGTGCTCGAGGCCCCAAGAGAGAGGCG 2391  
 QY 498 ThrTyrProAsnGlnProGlnAlaLysLeuTyrPheArgGlyIleAlaLysIleLeuSer 517  
 Db 2392 -----GAGCGCGTGGCGCGCGCTGCGCAAGAGAGATCATGAGAGGGGTGATTC 2439  
 QY 518 Leu-----LeuProIleGlu 522  
 Db 2440 CTGGCGTGGCCCTGGAG 2457  
 RESULT 16  
 US-08-481-238-21  
 : Sequence 21, Application US/08481238  
 : Patent No. 5795763  
 : GENERAL INFORMATION:  
 : APPLICANT: DAHLBERG, JAMES E.  
 : APPLICANT: LYAMICHEV, VICTOR I.  
 : APPLICANT: BROM, MARY ANN D.  
 : TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA  
 : NUMBER OF SEQUENCES: 29  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: MEDLEN & CARROLL  
 : STREET: 220 MONTGOMERY STREET, SUITE 2200  
 : CITY: SAN FRANCISCO  
 : STATE: CALIFORNIA  
 : COUNTRY: USA  
 : ZIP: 94104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patentin Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/481,238  
 : FILING DATE:  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:





Db	2215	AAGAGCTGGAGGAGGCGCGGAGCC	---	ATGAGCTTCACATGAGCTCCGGGACAC	2271
		:::			
Qy	461	ILleAluclueImetgluArghThrLeuAlaaspThrSerPro	-----	gluYrYla	477
		:::	:::		
Db	2272	GCGGCGACCTCATGAAGCTGTATGTGAlaAGCTTTCCACAGCTGAGGAAATGGGG			2331
Qy	478	TyrArValThrLeuAspArgHisAsnArgLeuIotrPhisAspProAlaThrArgLys			497
		:::		:::	
Db	2332	GCGGAGATGCTCTTCATGAGTCCAGCAGAGAGCTGTCTCTGAGGCGCCCAAAAGAGGGCG			2391
Qy	498	ThrTyrProAsnGluProGluAlaLysLleThrTyrArgIAlaAlaLysLlleuSer			517
		:::	:::		
Db	2392	-----GAGGCGCTGGCGCGCGCTGGCCAAAGAGATCATGAGGGGTGTATATCC			2433
Qy	518	Leu---LeuProIleGlu			522
Db	2440	CTGGCGCTGCCCTGTGAG			2457

```

Db 1741 -----ACCTCCAGACATCCCTCCGACCCCGCTTGGGAGAGATCCGCGGCC 1794
Qy 330 LeuylsGlnProgluLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTATCCGCGAGAGAGGGGTGCTATTGGCCCTGAGCTATAGCCAGATAGAGCTCAG 1854
Qy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyLeuasp 363
Db 1855 GTGCTGCCACCTCTCCGAGAGAACTGATCCGGGCTCTCCAGAGAGGGGGGAC 1914
Qy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspAla 377
Db 1915 ATCCACAGAGAGACCCGACGTGATGTTCGGCGTCCCGGAGAGCCGTGAC----- 1968
Qy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGGCGCGGGGCGGCGGACCAAGAC 1995
Qy 394 -----GlyLeuLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTGGGGTCTCTACGCGATGTGGCCACCGCTCTCCAGAGACTAGCCATC 2055
Qy 407 Pro----- 407
Db 2056 CCTTACAGAGAGCGCCAGCCTTCATTGAGCGCTACTTTCAGAGCTTCCCAAGTGGCG 2115
Qy 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCCTGATTTGAAAGACCTCTGAGAGCGAGAGCGCGGCGTACGTACGAGACCTCTTC 2175
Qy 424 AlaLysThrPheLeuValAspGlyLysArgGlnPheLeuGlySerPheAsnLeuAspPro 443
Db 2176 -----GCCCGCGCGCTACGTCAGCCAGACCTTGAAGCGCGGCTG 2214
Qy 444 ArgSerAlaArgLeuAsnThrGlnMetGlyValValLeuGlnSerPro-----Lys 460
Db 2215 AAGAGCGCGGAGAGCGCGCGAGCGC---ATGGCGCTCAACATGCGCGCGGAGCAC 2271
Qy 461 IleAlaGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCCCGCGCGCTCAATGAACCTGCTGAGCTCTTCCAGCGTGGAGCAATGGCG 2331
Qy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
Db 2332 GCCAGGATGCTCTTCAAGTCCAGAGAGCTGCTCTCGAGCGCCCAAGAGAGGGCG 2391
Qy 498 ThrTyrProAsnGlnProGluAlaLysLeuThrLysArgIleAlaAlaLysIleLeuSer 517
Db 2392 -----GAGGCGGTGGCGCGCGCTGGCCAAAGAGATCATGAGAGGGGTATGCC 2439
Qy 518 Leu---LeuProIleGlu 522
Db 2440 CTGGCGCTGCGCGCTGAG 2457

```

RESULT 19  
US-08-757-653-21

Sequence 21, Application US/08757653  
Patent No. 5843669  
GENERAL INFORMATION:  
APPLICANT: Kaiser, Michael W.  
APPLICANT: Lyamichev, Victor I.  
APPLICANT: Lyamichev, Natasha  
TITLE OF INVENTION: Cleavage of Nucleic Acid Using  
NUMBER OF SEQUENCES: 190  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/757,653  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02565  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2502 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-757-653-21

Alignment Scores:  
Pred. No.: 0.041 Length: 2502  
Score: 106.00 Matches: 119  
Percent Similarity: 33.70% Conservative: 65  
Best Local Similarity: 21.79% Mismatches: 174  
Query Match: 3.91% Indels: 188  
DB: 2 Gaps: 30

US-10-066-551-4 (1-525) x US-08-757-653-21 (1-2502)

```

Qy 118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal----- 130
Db 961 GATCTTCTGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 1020
Qy 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAACCCCTCAGAGACCTGAAGAGAGCGCGGCGCTCTCCGCAAAAGACCTGAGCTCTG 1080
Qy 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGAGCGCTTGGCTCCGCGCGCGAGACCCCATGTC----- 1128
Qy 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTGCGCTACTCTCTGGACCTTTCACACACCACCCCGAG 1167
Qy 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGGCGCGCGCTTACGCGGGGAGTGAAGAGAGAGGGGAGGGGCGCGCTT 1227
Qy 198 -----LeuGlyGlyArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGGCTTTCGCCAACCTGTGGGAGGCTTGAAGGGGAGAGAGGCTTCTTGG 1287
Qy 207 ---TyrPheLysValGlyLysLys-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACCGGAGGTGGAGAGAGCGCCCTTCCGCTCTCTGGCCACATGGAG-----GCC 1341
Qy 224 ThrGly-----SerValValGlyLys 230
Db 1342 ACGGGGGTGGCGCTGAGACGTGGCTATTCAGGCGCTTTCCTCGAGAGTGGCGGGGAG 1401
Qy 231 ValSer-----HisAspPheAsnArgTyrTrp 239
Db 1402 ATCCCGCGCTCGAGCGGAGGTCTCCGCTGCGCGCGCACCCCTTCAAC----- 1452
Qy 240 AlaSerHisSerAlaHisAsnAlaThrArgIleLeuArgSerGlyAsnIleGlyLysGly 259
Db 1453 ---CTCACTCCCGGAGACCACTGGAAAGGTCCTCTTGGAC-----GAGCTAAGG 1500

```



```

Qy 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCCCGCATCGCAAGACGAGAGACCGCAAGCCCTCCACAGCGCCCTCTG 1560
Qy 274 -----LeuLeuArgTyrArgGluThrVal 281
Db 1561 GAGGCGCTCCGAGGCGCCACCCATCGTGAGAGATCTCGATACCGGAGCTCAC 1620
Qy 282 GluGln-----SerProLeuTyrGlnIle-----GlnThrIleArg 294
Db 1621 AAGCTGAAGACACTTACATTGACCTTGGCGGACCTACACCCAGAGAGCGCGC 1680
Qy 295 IleAspTyrGlnSerValGlnThr-----ArgLeuIleSerAspPro 309
Db 1681 CTCACACCGCGCTTCACACACAGCGCCACGCGCAGCGGCTAAGAGCTCCATCC 1740
Qy 310 AlaIleGlyLeuAspArgAspArgIleGlyProProlIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCCAGAACATCCCGCTCCGACACCCCGCTTGGGACAGAGATCCGCGGCGC 1794
Qy 330 LeuIleGlnProGluIleGlySerValTyrIleValIleSerProTyr----- 343
Db 1795 TTCATCGCCGAGAGAGGCTGCTATGCTGCGCTGACCTATAGCCAGATAGAGCTCAGG 1854
Qy 344 PheValProThrIleGlySerGlyTyrAspAlaLeuAlaIleValGlnAspGlyIleAsp 363
Db 1855 GTGCTGGCGCCACCTCTCGCGGAGAGACCTGATCCGCGCTTCCAGAGAGCGGCGGAC 1914
Qy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACACGAGAGCGCGGCTGATGTTCCGCGCTCCCGCGGAGCGCGTGAC----- 1968
Qy 378 AlaValHisSerGlyTyrValIleGlyTyrArgIleGlyProLeuIleValAla----- 393
Db 1969 -----CCCTGATGCGCGCGGCGCGGCGCCAGAC 1995
Qy 394 -----GlyIleLeuIleGlyTyrGluGlnProAsnHisAlaVal 406
Db 1996 ATCAACTCGGCGGCTCTACGCGGATGTCGCGCCACCGCTCTCCAGAGAGCTACCATC 2055
Qy 407 Pro----- 407
Db 2056 CTTTACGAGAGAGCGCGGCTTATGAGCGCTACTTACAGAGCTTCCCAAGGTGCGG 2115
Qy 408 -----AlaThrIleAspIleGlyLeuThrIleGlySerValThrSerLeuHis 423
Db 2116 GCCTGATGAGAGAGACCTGAGAGAGCGGCGGCTGAGAGACCTCTTC 2175
Qy 424 AlaIleThrPheIleValAspGlyLeuArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GCGCGCGCGCTGCTGCGGAGAGAGCTGAGAGCGCGGCGT 2214
Qy 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerPro-----Lys 460
Db 2215 AAGACCGGTCGAGGCGCGGCGGAGCGC--ATGCGCTTCAACATCCCGCTCGGCGCAC 2271
Qy 461 IleAlaGlnGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCGCGCGAGCTCATGAGCTGCTGTAAGCTTTCGCCAGGCTGAGAGAAATGCGG 2331
Qy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
Db 2332 GCCAGGATGCTCTTCAGGTCACAGAGAGCTGCTCGAGAGCGCCCAAGAGAGGCGG 2391
Qy 498 ThrTyrProAsnGluProGluAlaIleLysLeuTyrIleValAlaIleLysIleLeuSer 517
Db 2392 -----GAGCGCGTGGCGCGCTGCGCAGAGAGCTATGAGAGGCTGATCC 2439
Qy 518 Leu---LeuProIleGlu 522
Db 2440 CTGCGCGTGGCGCGCTGAG 2457

```

```

RESULT 20
US-08-599-491-21
Sequence 21 Application US/08599491
Patent No. 5646717
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
APPLICANT: GROTELUESCHEN HALL, JEFF S.
APPLICANT: LYAMICHEV, VICTOR
APPLICANT: OLIVE, DAVID M.
APPLICANT: PRUDENT, JAMES R.
TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
NUMBER OF INVENTION: INVADER-DIRECTED CLEAVAGE
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk.
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,491
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: P-40,027
REFERENCE/DOCKET NUMBER: FORS-01802
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-599-491-21
Alignment Scores:
Pred. No.: 0.041 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
DB: 2 Gaps: 30
US-10-066-551-4 (1-525) x US-08-599-491-21 (1-2502)
Qy 118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal----- 130
Db 961 GATCTTCTGCGCCCTGCGCGCCGCGGAGGCGGCTCCACCGGCGCCGAGCTTAT 1020
Qy 131 ArgLeuLeuLeuAspAsnAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAAGCCCTCAGGAGCTGAGAGAGCGGCGGCTTTCGCCAAGAGCTGAGCGTTCG 1080
Qy 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGAGCGCTTGGCTCCCGCGCGGCGAGACCCATGCTC----- 1128
Qy 164 ValLeuArgLysTrpArgAlaLeuGlyTyrIleuThrAsp-----PheProArg 179
Db 1129 -----CTCGCTTACCTCTCGAGACCTTCACACACACCGCCGAG 1167
Qy 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197

```

```

Db 1168 GGGGTGCGCCGCGCTACGCGGGGAGTGCAGCGAGCGGGGGAGCGCGCCCTT 1227
QY 198 -----LeuGlyIYArGAsnIleGlyAspGlu-----Lys 206
Db 1228 TCCGAGAGGCTCTTCCGCAACCTGTGGGAGAGGCTTGAGGGGAGAGAGAGCTCTTGG 1287
QY 207 ---TyrPheIysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACCGGAGGTGCAGAGGCGCCCTTCCGCTGTCTGCGCCACATGAG-----GCC 1341
QY 224 ThrGly-----SerValValGlyGlu 230
Db 1342 ACGGGGAGGCGCCCTGAGAGCTGCTATCTCAGAGCCCTTGTCCCTGAGGTGGCGGGAG 1401
QY 231 ValSer-----HisAspPheAspArgTyrTrp 239
Db 1402 ATCCCGCGCTCGAGCGCGAGGTCTTCCGCGCGCGCCACCCCTTCAAC----- 1452
QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259
Db 1453 ---CTCAACTCCCGGAGCAGCTGTGAAAGGGTCTCTTTGAC-----GAGCTAAGG 1500
QY 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCGCGCATCGCGAGAGCGAGAGACCGGCAAGCGCTCCAGCAGCGCGCTCTTG 1560
QY 274 -----LeuLeuArgTyrArgGluThrVal 281
Db 1561 GAGCCCTCCGCGAGGCCACCCCATGCTGTGAGAAATCCGAGTACCGGAGCTCAC 1620
QY 282 GluGln-----SerProLeuTyrGlnIysIle-----GlnThrGlyArg 294
Db 1621 AAGGTGAAGAGCACCTTACATTGACCCCTTGGCGGAGCTCATCCACCCAGAGCGGGCGC 1680
QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309
Db 1681 CTCACACCGCGCTTCAACCAAGCGGCGAGCGGCGAGGCTAAGAGCTCCGATCC 1740
QY 310 AlaIysGlyLeuAspArgAspArgArgLysProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 ---AACCTCCAGAACATCCCGCTCCGCGACCCGCTTGGGAGAGGATCCGCGGGCC 1794
QY 330 LeuLysGlnProGlyLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCAATCCCGGAGGGGTGGCTATTGGTGGCCCTGAGCTATTAACCAAGTAGACTCAGG 1854
QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysValGlnAspGlyIleAsp 363
Db 1855 GTGCTGGCGCCACTCTCCGCGAGAGAACTGATCCGGGTCTTCCAGAGAGCGGGGAGC 1914
QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaIleThrAspValAla 377
Db 1915 ATCCACAGCGAGCGCGAGCTGTGATGTTCGCCGCCGCGGAGCGCTTGAC----- 1968
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGGCGCGGGCGGCCAGAGACC 1995
QY 394 -----GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTGGGGTCTCTAGAGCATGTGGCGCCCTCTCCAGAGAGCTAGCCATC 2055
QY 407 Pro----- 407
Db 2056 CCTTACAGAGAGCGCCGCTTATTAGCGCTTTCAGAGCTTCCCAAGGTGCGG 2115
QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCCTGATTTAGAGAACCTTGGAGAGGAGGCAAGAGCGGGGTACGTGGAGACCTCTTC 2175
QY 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GGCGCGCGCGCTAGCTGTGCAGACCTAAGAGCGCGGGGTG 2214

```

```

QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGlnSerPro-----Lys 460
Db 2215 AAGAGCGTGGGAGGCGCGCGAGCGC---ATGGCTTCAACATGCCCCGTCCGGGAGC 2271
QY 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCCCGCGCATCAAGAGCTGATGTGTGAAGCTTTCGCCAGCTGAGAGAAATGGGG 2331
QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheIleAspProAlaThrArgLys 497
Db 2332 GCCAGATGCTCTTCAAGTCCAGAGAGAGCTGTCTCTGAGGCGCCAAAAGAGAGGCG 2391
QY 498 ThrTyrProAsnGluProGluAlaLysLeuThrPyrAspArgIleAlaIleLysIleLeuSer 517
Db 2392 -----GAGCGCGTGGCGCGCTGCGCCAGAGAGTCAAGAGGAGGTGTATCCC 2439
QY 518 Leu---LeuProIleGlu 522
Db 2440 CTGGCGCGTGGCGCGCTGAG 2457

RESULT 21
US-08-756-386-21
; Sequence 21, Application US/08756386
; Patent No. 5985557
; GENERAL INFORMATION:
; APPLICANT: Prudent, James R.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States Of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,386
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/682,853
; FILING DATE: 12-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,491
; FILING DATE: 24-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ingolia, Diane E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: FORS-02564.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ. ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-756-386-21

Alignment Scores:
Pred. No.: 0.041 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65

```

Best Local Similarity:	21.79%	Mismatches:	174
Query Match:	3.91%	Indels:	188
DB:	2	Gaps:	30

US-10-066-551-4 (1-525) x US-08-756-386-21 (1-2502)

Oy	118	AsnLeuValIYrLeuAlaAlaGluArgGlyValArgVal-----	130	
Db	961	GATCTTGTGGCCCTCGAGCCGCCACAGGGGGGGCGSGGTCCACCGGGCCCCGAGCCTTAT	1020	
Oy	131	ArgLeuLeuLeuAspAspAsnaAnThrArgGlyLeu-----	AspAspLeuLeu 147	
Db	1021	AAACCCCTCAGGAGACCCTGAAGGAGCGGGGGGCTTCGCACAAGACTGAGCCTTG	1080	
Oy	148	AlaLeuAspSer-----HisProAsnIlleGluValArgLeuPheAsnProPhe	163	
Db	1081	GCCCTTAGGGAGAAGCCCTTGCCCTCCGCCCGCGAGACCCCATGCTC-----	1128	
Oy	164	ValLeuArgLyStrParGalAlaLeuGlyTyrrLeuThrAsp-----	PheProArg 179	
Db	1129	-----CTCGGCTTACCTCTCGAGACCCCTCCAACACCAACCCCGAG	1167	
Oy	180	LeuAsnArgArgMetHisAsnLysSerPheThr-----	AlaAspAsnArgAlaThrIle 197	
Db	1168	GGGTGTGGCCCGCGCTACGCGGGGAGTGGAGAGAGAGCGGGGAGCGGCCCTT	1227	
Oy	198	-----LeuGlyGlyArgAsnIlleGlyAspGlu-----	206	
Db	1228	TTCGAGAGGCTTTCCGCAACCTGTGGGGAGGCTTGAAGGGGAGAGAGGCTCTTTGG	1287	
Oy	207	--"TyrrPheLysValIglYgu-----	AspThrValPheAlaAspLeuAspIlleLeuAla 223	
Db	1288	CTTTACCGGGAGGTGGAGAGAGCCCTTTCCGCTGTCTGTGCCACATGAG-----	GCC 1341	
Oy	224	ThrsIy-----	SerValValcIyGu 230	
Db	1342	ACGGGGGTGGCGCTCGACGTGCGCTATTCTCAGGGCCTTGCTCCCTGAGAGTGGCCGGGAG	1401	
Oy	231	ValSer-----	HisAspPheAspArgTrpTrp 239	
Db	1402	ATCGCCCGCCTCGAGCCGAGGCTTCCGCGCTGGCGGCGCACCCCTTAC-----	1452	
Oy	240	AlaSerHisSerAlaHisAsnaIaThrArgIleileArgSerGlyAsnIlleolyLysGly	259	
Db	1453	---CTCAACTCCGGGAGCACAGCTGGAAGAGGTCCTTTGAC-----	GAGCTAAGG 1500	
Oy	260	LeuGlnAlaLeuGlyTyrrAsnAspGluThrSerArgHisAla-----	273	
Db	1501	CTTCCCGCATCGGCAAGACGAGAGAACCGGCACACCGCTCCACGAGCGCGCTCTG	1560	
Oy	274	-----	LeuLeuArgTyrrArgGluThrVal 281	
Db	1561	GAGGCCCTCGCGAGGCCACCCACCCATCGTGGAGAMAGATCTTCACTACCGSAGCTCAC	1620	
Oy	282	GluIn-----	SerProLeuTyrrGlnLysIle-----	GlnThrGlyArg 294
Db	1621	AAGCTGAAGAGACACTFACATTGACCCCTTGGCGAGCTCATCCACCCAGACGGGGCCG	1680	
Oy	295	IleAspTrpGlnSerValGlnThr-----	ArgLeuIlleSerAspAspPro 309	
Db	1681	CTCCACACCCGCTTCAACACGAGCGCCACGCGCCAGCGGAGGCTTAAGTCTCCCATCCC	1740	
Oy	310	AlaLysGlyLeuAspArgAspArgLysProProIlleAlaGlyArgGlnAspAla	329	
Db	1741	-----AACTCCAGAAACATCCCGCTCCGACCCGCTTGGGCAAGAGATCCCGCGGCC	1794	
Oy	330	LeuLysGlnProGluLysSerValTyrrLeuValSerProTyrr-----	343	
Db	1795	TTTCATCCGCCAGAGAGGGGTGGCTATTGGTGCGCTTGAGATTATAGCCAGATGAGCTCAG	1854	
Oy	344	PheValProThrIlysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyLysP	363	
Db	1855	GTCGTGGCCCACTCTCTCGCGCAGCGAAACTGTATCCCGGGCTTCTCAGAGAGGGGAG	1914	

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-823-516-21

Alignment Scores:
Pred. No.: 0.041 length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
DB: 2 Gaps: 30

US-10-066-551-4 (1-525) x US-08-823-516-21 (1-2502)
QY 118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal-----130
Db 961 GANCTTCTGGCCCTGCGCGCGCCGAGGGGGCGGCTCCACCGCGCCCGACCTTAT 1020
QY 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAAGCCCTCAGGGACCTGAAGAGAGCGGGGGCTCTCCGCAAAAGACCTGAGGCTTCG 1080
QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGAGGCTTGGCTCCGCCCGCAGCAGACCCCATGCTC-----1128
QY 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTGCGCTACCTCCTGAGACCTTCGACACCTGCAACACCCCGCGAG 1167
QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGGCCGCGGCTACGGCGGGAGTGGACGAGAGAGAGGCGGGCGCGCTT 1227
QY 198 -----LeuGlyGlyArgAsnIleGlyAspGlu-----206
Db 1228 TCCGAGAGGCTCTTCCCAACCTGTGGGAGGCTTGAGGGGGAGAGAGGCTCTTGG 1287
QY 207 ---TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACGGGAGGTGGAGAGGCGCCCTTCCGCTGTCTCGCGCCACATGAG-----GCC 1341
QY 224 ThrGly-----SerValValGlyGlu 230

```

```

Db 1342 ACGGGGTGGCGCTGAGAGTGGCTATCTCAGGCGCTTGTCCCTGAGAGTGGCGGGAG 1401
QY 231 ValSer-----HisAspPheAspArgTyrTrp 239
Db 1402 ATCGCCCGCCCTCGAGCGCGAGGCTTCCGCTGGCCCGCCACCCCTTCAAC-----1452
QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGly 259
Db 1453 ---CTCACTTCGCGGAGCCAGCGTGGAAAGGGTCTCTTTCAC-----GAGCTAGGG 1500
QY 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla-----273
Db 1501 CTTCCCGCATCGCGCAAGAGAGAGAACCGGAGCGCTCCACACAGCGCGCGCTCTG 1560
QY 274 -----LeuLeuArgTyrArgGluThrVal 281
Db 1561 GAGGCGCTCCGCGAGGCCACCCCATGCTGAGAGAGATCTCGAGTACCGGGAGCTCACC 1620
QY 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294
Db 1621 AAGCTGAAGAGACACTTACATTGACCCCTTGGCGGACCTACCCAGAGAGCGGGCGG 1680
QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309
Db 1681 CTCACACCGCGCTTCAACACAGAGCGGCGACGCGGCGGAGGCTTAAGTACTCGATCCC 1740
QY 310 AlaLysGlyLeuAspArgAspArgArgLysProPheIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACTCCAGAACATCCCGCTCCGACACCGCGCTTGGCAGAGATCCCGCGGCC 1794
QY 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr-----343
Db 1795 TTCATCGCCGAGAGGGGGTGGTATGTTGGCTGAGCTATGACCGAGATAGAGCTCAGG 1854
QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363
Db 1855 GTGCTGGCCACCTCTCCGGGAGAGACCTGATCCGGCTTTCAGAGAGGGCGGGAG 1914
QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACACGGAGACCGGCGACGTGATGTGGCGTCCCGCGGAGCGCGTGGAC-----1968
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla-----393
Db 1969 -----CCCTGTATGGCGCGCGCGGCCACMACC 1995
QY 394 -----GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406
Db 1996 ATCACTTCGGGGTCTCTAGGGCATGTGGCCACCGCTCTCCAGAGCTAAGCCATC 2055
QY 407 Pro-----407
Db 2056 CCTTACGAGAGAGCGCCAGCGCTTCATTGAGCGCTTTCAGAGCTTCCCAAGGTGCGG 2115
QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerValThrSerLeuHis 423
Db 2116 GCCTGATTTAGAGAGACCTGAGAGAGGAGGCGGGGTACGTGAGGACCTCTTC 2175
QY 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GCGCGCGCGCTACGTGCGAGACCTTGAAGGCCCGGGGTG 2214
QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerPro-----Lys 460
Db 2215 AAGAGCTGCGGAGGGCGCGGAGCGC---ATGCGCTTCAACATGCGCGCGGGGAGC 2271
QY 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCCCGCAGACCTTGAAGAGTGGATAGAGCTTCCCGCAGAGCTGAGAAATGGGG 2331
QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497

```

Db 2332 GCCAGATGCTCTTCAGTCCACGACGAGCTGTCTCGAGGCCCAAGAGAGGCG 2391  
 Oy 498 ThrlyProAsngluProgluAlaLysleuTrpLysArgIleAlaLysIleleuSer 517  
 Db 2392 -----GAGCCGTGGCCCGGCTGGCCAGAGAGTCAAGAGGGGTATCC 2439  
 Oy 518 Leu---LeuProIleGlu 522  
 Db 2440 CTGGCCGTGGCCCTGGAG 2457  
 RESULT 23  
 US-08-682-853A-21  
 : Sequence 21, Application US/08682853A  
 : Patent No. 6001567  
 : GENERAL INFORMATION:  
 : APPLICANT: Brow, Mary Ann D.  
 : APPLICANT: Groelueschen Hall, Jeff S.  
 : APPLICANT: Lyamichev, Victor  
 : APPLICANT: Olive, David M.  
 : APPLICANT: Prudent, James R.  
 : TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY  
 : TITLE OF INVENTION: INVADER-DIRECTED CLEAVAGE  
 : NUMBER OF SEQUENCES: 69  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Medlen & Carroll, LLP  
 : STREET: 220 Montgomery Street, Suite 2200  
 : CITY: San Francisco  
 : STATE: California  
 : COUNTRY: United States Of America  
 : ZIP: 94104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/682,853A  
 : FILING DATE: 12-JUL-1996  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 08/599,491  
 : FILING DATE: 24-JAN-1996  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Ingolia, Diane E.  
 : REGISTRATION NUMBER: 40,027  
 : REFERENCE/DOCKET NUMBER: FORS-02306  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (415) 705-8410  
 : TELEFAX: (415) 397-8338  
 : INFORMATION FOR SEQ ID NO: 21:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2502 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: double  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: DNA (genomic)  
 : US-08-682-853A-21  
 Alignment Scores:  
 Pred. No.: 0.041 Length: 2502  
 Score: 106.00 Matches: 119  
 Percent Similarity: 33.70% Conservative: 65  
 Best Local Similarity: 21.79% Mismatches: 174  
 Query Match: 3.91% Indels: 188  
 DB: Gaps: 30  
 US-10-066-551-4 (1-525) x US-08-682-853A-21 (1-2502)  
 Oy 118 AsnleuValIlyrleuAlaAlaIargIlyValArgVal----- 130  
 Db 961 GATCTTCGGCCCTGGCCGCGCCAGGGGGGCGGTCACCGGGCCCGAGCCTTAT 1020  
 Oy 131 ArgleuLeuLeuAspAspAsnThrArgIlyLeu-----AspAspLeuLeu 147

Db 1021 AAAGCCCTCAGGAGACCTGAGAGGCGGGGCTTCGCCAAGACCTGAGCTTCTG 1080  
 Oy 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163  
 Db 1081 GCCCTGAGGGAAGGCTTGGCTCCGCCGCGCAGACCCCATGCTC----- 1128  
 Oy 164 ValLeuArgLysTrpArgAlaLeuGlyTrpLeuThrAsp-----PheProArg 179  
 Db 1129 -----CTGCTTACTCTCTGGACCTTCACACACACCCCGAG 1167  
 Oy 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197  
 Db 1168 GGGGTGGCCCGGCGCTACGGGGGAGTGAACGAGAGGAGGAGGCGGGCCGCTT 1227  
 Oy 198 -----LeuGlyIlyArgAsnIleGlyAspGlu----- 206  
 Db 1228 TCCGAGAGCTCTTCGCCACCTGTGGGGGAGGCTTGAGGGGAGAGAGCTCCTTGG 1287  
 Oy 207 ---TyrPheLysValIlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223  
 Db 1288 CTTTACCGGAGGTGAGAGAGCCCTTTCGCTGTCTGGCCACATGAG-----GCC 1341  
 Oy 224 ThrGly-----SerValValIlyGlu 230  
 Db 1342 ACGGGGTGGCGCTGTGAGCTGAGCTATCTCAGGCTTGTCCCTGAGAGTGGCGGGAG 1401  
 Oy 231 ValSer-----HisAspPheAspArgIlyTrp 239  
 Db 1402 ATCCGCCCTCGAGCGGAGCTTCCGCTGGCGGGCCACCCCTTCAC----- 1452  
 Oy 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259  
 Db 1453 ---CTCACTCCCGGAGCAGCTGSAAGGGTCTCTTTCAC-----GAGCTAAGG 1500  
 Oy 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273  
 Db 1501 CTTCGCCGATCGCAGACGAGAGAGACCGGACGCTCCACAGCGCGCGCTCTG 1560  
 Oy 274 -----LeuLeuArgTyrArgIlyThrVal 281  
 Db 1561 GAGGCGCTCCGAGCGCCACCCCATCTCGTGAGAGATCTTCATACCGGGAGCTCAC 1620  
 Oy 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294  
 Db 1621 AAGCTGAAGACACCTTACATTGACCCCTTCCGCGACTTCATCCACCCAGAGCGGCGC 1680  
 Oy 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309  
 Db 1681 CTCACACCCCGCTTCACACGACGCGCACGCGGACGAGGAGCTAAGTACGCGCATCC 1740  
 Oy 310 AlaLysGlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAla 329  
 Db 1741 -----AACCTCCAGAACATCCCGCTCCGACCCCGCTTGGGCAAGATCCGCGGGCC 1794  
 Oy 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343  
 Db 1795 TTCATGCCGAGAGAGGCTGCTATGTGGTGGCTGACCTAAGCCAGATAGACTCAGG 1854  
 Oy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspIlyIleAsp 363  
 Db 1855 GTGCTGGCCCACTCTCCGGGAGACGACGACGACGACGAGGAGGAGGCGGAGAC 1914  
 Oy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377  
 Db 1915 ATCCACAGGAGACCGCAGCTGTGATGTCGCGCTCCCGCGGAGCGCTGAGAC----- 1968  
 Oy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393  
 Db 1969 -----CCCTGATGCGCGGGGCGGCGGCAAGACC 1995  
 Oy 394 -----GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406

```

Db 1996 ATCAACTGGGGGCTCTCTACAGCATATGTGGCCACCGCCTCTCCAGAGACTACCATC 2055
Oy 407 Pro----- 407
    |||
Db 2056 CCTTACGAGAGAGGCCAGGCTTCATTGAGCGCTACTTTCAAGACTTCCCAAGGTGGG 2115
Oy 408 -----AlaThrLysAspLysGlyLeuThrGlySerValThrSerLeuHis 423
    |||
Db 2116 GCCTGATTTGGAAGACCTCTTGAGGAGGAGGAGGAGGGGGGTACTGTGAGACCTCTTC 2175
Oy 424 AlAlysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
    |||
Db 2176 -----GGCCGGCGCCGCTACTGTCCAGACCTAGAGGCCGGGTG 2214
Oy 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerPro-----Lys 460
    |||
Db 2215 AAGAGCGTGGCGGAGGGCGCCAGCCG---ATGCGCTTCAACATGCGCTCCGGGACAC 2271
Oy 461 IleAlaGluGlnMetGluArgThrIleuAlaAspThrSerPro-----GluTyrAla 477
    |||
Db 2272 GCCCCGACCCCAAGCAAGCTGCGTGAAGCTTCTCCAGCGCTGAGAGAAATGGGG 2331
Oy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
    |||
Db 2332 GCCAGGATGCTCTTCAAGGTCCACGACGACTGTCTCTGAGGCCCAAAAGAGAGGGCG 2391
Oy 498 ThrTyrProAsnGlnProGluAlaLysLeuThrPysArgIleAlaLysIleLeuSer 517
    |||
Db 2392 -----GAGCGCGTGGCGCCGCTGCGCAAGAGATGATGAGGAGGCTGATTC 2439
Oy 518 Leu---LeuProIleGlu 522
    |||
Db 2440 CTGGCCGTCGCCCCGAG 2457

RESULT 24
US-08-759-038-21
Sequence 21, Application US/08759038
Patent No. 6090543
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Dahlberg, James E.
TITLE OF INVENTION: Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,038
FILING DATE: 02-DEC-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.

```

```

:      REGISTRATION NUMBER: 40,027
:      REFERENCE/DOCKET NUMBER: FOS-02574
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: (415) 705-8410
:      TELEFAX: (415) 397-8338
:      INFORMATION FOR SEQ ID NO: 21:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 2502 base pairs
:      TYPE: nucleic acid
:      STRANDEDNESS: double
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      US-08-759-038-21

Alignment Scores:
Pred. No.:      0.041      Length:      2502
Score:      106.00      Matches:      119
Percent Similarity:      33.70%      Conservative:      65
Best Local Similarity:      21.79%      Mismatches:      174
Query Match:      3.91%      Indels:      188
DB:      Gaps:      30

US-10-066-551-4 (1-525) x US-08-759-038-21 (1-2502)

OY      118      AsnLeuValTyrIleuAlaIaGluArgGlyValArgVal-----130
:      :::::      |||||      |||||      |||||
Db      961      GATCTTCTGGCCCTGGCCGCGCCGCGAGGGGGGGCGGTCCACCGGCCCCGAGCCTTAT      1020
:      :::::      |||||      |||||      |||||

OY      131      ArgLeuLeuLeuAspAspAsnAnThrArgGlyLeu-----AspAspLeuLeu      147
:      :::::      |||||      |||||      |||||
Db      1021      AAGCCCTCAGGACCTGAAAGAGAGCGCGGGGCTTCTGCCAAGACCTGAGCCTTG      1080
:      :::::      |||||      |||||      |||||

OY      148      AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe      163
:      |||||      |||||      |||||      |||||
Db      1081      GCCCTGAGGAGGAGCCTTGCGCTCCGCGCGGAGGAGACCCCATGCTC-----      1128
:      :::::      |||||      |||||      |||||

OY      164      ValLeuArgLysTrpArgAlaLeuGlyTyrIleThrAsp-----PheProArg      179
:      |||||      |||||      |||||      |||||
Db      1129      -----CTCGCCTACCTCTCGAGCCCTTCCACACACACCCCGAG      1167
:      :::::      |||||      |||||      |||||

OY      180      LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle      197
:      |||||      |||||      |||||      |||||
Db      1168      GGGGGGCGCGCGCCTACGCGGGGGAGTGAAGAGAGAGCGGGGAGACGGCGCCCTT      1227
:      :::::      |||||      |||||      |||||

OY      198      -----LeuGlyGlyArgAsnIleGlyAspGlu-----206
:      :::::      |||||      |||||      |||||
Db      1228      TCCGAGAGGCTCTTCGCCAACCTGTGGGGAGGCTGAGGGGAGAGAGGCTCTTGG      1287
:      :::::      |||||      |||||      |||||

OY      207      --TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla      223
:      |||||      |||||      |||||      |||||
Db      1288      CTTTACGGGAGGAGTGGAGAGGCGCCCTTCCGCTGCTGTCGCCACATGAG-----GCC      1341
:      :::::      |||||      |||||      |||||

OY      224      ThrGly-----SerValValGlyGlu      230
:      |||||      |||||      |||||      |||||
Db      1342      ACGGGGGGCGCCTGAGACGTGGCCTATCTTCAGGGCCTTGTCCTGAGAGTGGCCGGGAG      1401
:      :::::      |||||      |||||      |||||

OY      231      ValSer-----HisAspPheAspArgTyrTrp      239
:      :::::      |||||      |||||      |||||
Db      1402      ATCGCCGCCCTCGAGGCCCGAGGCTTCCGCGCTGCGCGGCCACCCCTTCAAC-----      1452
:      :::::      |||||      |||||      |||||

OY      240      AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly      259
:      :::::      |||||      |||||      |||||
Db      1453      ---CTCAACTCCCGGAGCACGCTGMAAAGGTCCTCTTGAC-----GAGCTAAGG      1500
:      :::::      |||||      |||||      |||||

OY      260      LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla-----273
:      |||||      |||||      |||||      |||||
Db      1501      CTTCCCGCATCGCGCAGAGAGAGAACCGGACCGCTCCACACAGCGCGCGCTCTTG      1560
:      :::::      |||||      |||||      |||||

OY      274      -----LeuLeuArgTyrArgGluThrVal      281
:      :::::      |||||      |||||      |||||
Db      1561      GAGCGCCTCGCGAGGAGGCCACCCCATCTGAGAGAGATCTCTCACTACCGGAGCTACC      1620
:      :::::      |||||      |||||      |||||

OY      282      GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg      294
:      :::::      |||||      |||||      |||||

```

```

Db 1621 AAGCTGAAGACGACCTACATGACCCCTTGCCGAGACCTCATCCACCCCGAGGCGCCG 1680
      ::::: ||||| ::::: |||||
Qy 295 lileasptpglnservalglnthr-----Argleulleasrpspro 309
      ::::: ||||| ||||| |||||
Db 1681 CTCACACACCCGCTTACACGAGGCGACGCGGCGGAGGCTAGTCCGATCCG 1740
      ::::: ||||| ||||| |||||
Qy 310 AlalysglYleuAspArgspArglysProProlleAlglYArgleuGlnAspAla 329
      ||| ||| ::::: ||||| |||||
Db 1741 -----AACCTCAGAACATCCCGCTCCGACCCCGCTTGGGACAGAGATCCGCGGCG 1794
      ||| ||| ::::: ||||| |||||
Qy 330 LeuYsglnProglYlysServalYtyrleuValSerProtyr----- 343
      ||| ||| ::::: ||||| |||||
Db 1795 TTCATCGCGCGAGAGGCGGCTATGAGGCGCTGACCTAGATAGCCAGATAGAGTCAAG 1854
      ||| ||| ::::: ||||| |||||
Qy 344 PheValProthYrlySserglYthrAspAlaAlaYleuValGlnAspGlylleasP 363
      ::::: ||||| ::::: ||||| |||||
Db 1855 GTGCTGGCGCCACCTCTCCGCGAGGAGAACCTGATCCGGTCTTCAGAGAGGCGGAGC 1914
      ::::: ||||| ::::: ||||| |||||
Qy 364 ValThrValleuThrAsnSer-----LeuGlnAlaThrAspValAla 377
      ||| ||| ||| |||
Db 1915 ATCCACACGAGACCGCCGACGCTGATGTCGGCGTCCCGGAGGCGCTGGAC----- 1968
      ||| ||| ||| |||
Qy 378 AlalalHisSerGlyTyryVallyStryArglySProleuLeuYsAla----- 393
      ||||| ::::: ||||| |||||
Db 1969 -----CCCTCATGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1995
      ||||| ::::: ||||| |||||
Qy 394 -----GlylleYsleuTyrglLeuGlnProAsnHisAlaVal 406
      ||||| ::::: ||||| |||||
Db 1996 ATCAACTTCGGGGCTCTTACGCGATGTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2055
      ||||| ::::: ||||| |||||
Qy 407 Pro----- 407
      |||
Db 2056 CTTTACGAGAGGCGGCGCTTACATGAGCGCTTACAGACCTTCCCAAGTCCG 2115
      |||
Qy 408 -----AlaThrYsAspYsglyleuThrGlySerServalThrSerleuHis 423
      ||| ||||| |||||
Db 2116 GCCTGATGAGAGACCTCGGAGGAGGCGGAGGCGGCGGAGTGGAGACCTCTTC 2175
      ||| ||||| |||||
Qy 424 AlalYsthrPheIleValAspGlylsArgIlePheIleGlySerPheAsnleuAspPro 443
      ||||| ||||| |||||
Db 2176 -----GCCCGCGCGCTACGTCACAGACCTAGAGCGCGCGGCTG 2214
      ||||| ||||| |||||
Qy 444 ArgSerAlaArgleuAsnThrGlnMetGlyValIleGlnSerPro-----Lys 460
      ::::: ||||| ||||| |||||
Db 2215 AAGAGGCTCGGAGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2271
      ::::: ||||| ||||| |||||
Qy 461 lileAlaGlnMetGlnArgThrleuAlaAspThrSerPro-----GluTyraAla 477
      ||||| ||||| |||||
Db 2272 GCGCGCGACCTCATGAAGCTGAGTGTGAAGCTCTCCCGGCGGCGGCGGCGGCGGCGG 2331
      ||||| ||||| |||||
Qy 478 TyrArgValThrleuAspArgHisAsnArgleuGlnThrPheAspProAlaThrArgly 497
      ||||| ||||| ||||| |||||
Db 2332 GCCAGATGCTCTTCAAGTTCACAGAGCTGCTCCGAGGCGGCGGCGGCGGCGGCGGCGG 2391
      ||||| ||||| ||||| |||||
Qy 498 ThrTyProAsnGlnProGlnAlaYsleuThrPlyArgIleAlaAlaYsIleleuSer 517
      ||| ||| ||||| |||||
Db 2392 -----GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2439
      ||| ||| ||||| |||||
Qy 518 Leu--LeuProIleGlu 522
      ||| ||||| |||||
Db 2440 CTGCGCGTGCCTCGAG 2457
      ||| ||||| |||||

```

```

RESULT 25
US-08-758-314-21
; Sequence 21, Application US/08758314
; Patent No. 6090606
; GENERAL INFORMATION:
; APPLICANT: Kaiser, Michael W.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Lyamichev, Natasha
; TITLE OF INVENTION: Improved Cleavage Agents
; NUMBER OF SEQUENCES: 134

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,314
FILING DATE: 02-DEC-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 29-NOV-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02575
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ. ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2502 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-758-314-21
Alignment Scores:
Score: 0.041 Length: 2502
Percent Similarity: 106.00 Matches: 119
Best Local Similarity: 33.70% Conservative: 65
Query Match: 21.79% Mismatches: 174
DB: 3 Indels: 188
Gaps: 30
US-10-066-551-4 (1-525) x US-08-758-314-21 (1-2502)
Qy 118 AsnleuValTyrleuAlaAlaGluArgGlyValArgVal----- 130
      ::::: ||||| ||||| |||||
Db 961 GATCTCTGGCCCTCGCGCGCGCGGCGGCGGCGGCTCCACCGGCGGCGGCGGCGGCGGCGG 1020
      ::::: ||||| ||||| |||||
Qy 131 ArgleuLeuAspAspAsnThrArgGlyleu-----AspAspleuLeuLeu 147
      ::::: ||||| ||||| |||||
Db 1021 AAGCCCTCAGGACCTGAGAGGAGGCGGCGGCGGCTTCGCCAAGACCTGACGCTTCG 1080
      ::::: ||||| ||||| |||||
Qy 148 AlaLeuAspSer-----HisProAsnIleGluValArgleuPheAsnProPhe 163
      ||||| ||||| |||||
Db 1081 GCCCTGAGGAGGAGGCTTGGCGTCCGCGCGGAGACCCCATGCTC----- 1128
      ||||| ||||| |||||
Qy 164 ValLeuArglystrPArgAlaLeuGlyTyrleuThrAsp-----PheProArg 179
      ||||| ||||| |||||
Db 1129 -----CTGGCTTACCTCTCTGACCTTCCACACACCGGCGGCGGCGGCGGCGGCGGCGG 1167
      ||||| ||||| |||||
Qy 180 LeuAsnArgArgMetHisAsnYsSerPheThr-----AlaAspAsnArgAlaThrIle 197
      ||| ||| ::::: ||||| |||||
Db 1168 GGGGTGGCGCGGCGCTACGCGGCGGAGTGAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1227
      ||| ||| ::::: ||||| |||||
Qy 198 -----LeuGlyGlyArgAsnIleGlyAspGlu----- 206
      ||| ||| ||||| |||||

```

```

Db 1228 TCCGAGAGGCTTCCGACACCTGTGGGGAGGCTTGAGGGGAGAGAGCTCTTGG 1287
Qy 207 --TyrPheValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTATCCGGAGGAGTGAAGAGGCCCTTCCGCTGTCCGCGCCACATGAG-----GCC 1341
Qy 224 ThrGly-----SerValAlaGlyGlu 230
Db 1342 ACGGGGGTCCCTGAGAGCTGCTCATCTCAGGGCTTGTCCCTGAGGTGCGCGGGAG 1401
Qy 231 ValSer-----HisAspPheAspArgTyrThr 239
Db 1402 ATCCGCCCTCGAGAGCGGAGGTCTCCGCTGCGCGCCACCTTCAAC----- 1452
Qy 240 AlasSerHisSerAlaHisAsnAlaThrArgTleIleArgSerGlyAsnIleGlyLysGly 259
Db 1453 --CTCAACTCCCGGAGACACTGGAAGGGTCTCTTTGAC-----GAGCTAAGG 1500
Qy 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCCCGCATCGCAGAGCAGAGAGACCGCAGCGCTCCACCGAGCGCGCTCTG 1560
Qy 274 -----LeuArgTyrArgGluThrVal 281
Db 1561 GAGGCCCTCCGAGAGGCCACCATCTGAGAGAGATCTGCACTACCGGAGCTCAC 1620
Qy 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294
Db 1621 AAGCTGAGAGACACTTACATTGACCTTGCCTGCGACCTGACCTACCCAGCGGGCGC 1680
Qy 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309
Db 1681 CTCACACCCGCTTACACAGAGCGGCACAGCGGCGGAGGCTAAGTACCTCCATCC 1740
Qy 310 AlaLysGlyLeuAspArgAspArgLysProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCCAGAACATCCCGCTCGCAGCCCGCTTGGCAGAGATCCCGCGGC 1794
Qy 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCAATCCGCGAGGGGTGCTATTGCTGCGCTGAGCTAGCCAGATAGAGCTCAG 1854
Qy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363
Db 1855 GTGCTGCCCACTCTCTCGGAGAGACGATCGCGGCTTCCAGAGAGGGGGGAGC 1914
Qy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACAGAGAGACCGGAGCTGATGTCGGCGTCCCGGAGAGCGCTGGAC----- 1968
Qy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGCGCGCGGCGGCGGCGCAAGAC 1995
Qy 394 -----GlyTleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTGGGGTCTCTACGCGATGCGCCACCGCTCTCCAGAGACTACCATC 2055
Qy 407 Pro----- 407
Db 2056 CCTTACGAGAGAGCGCCCTTCAATTAGCGCTTTCAGAGCTTCCCAAGGTGGG 2115
Qy 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCTGTGATTGAGAGACCTGAGAGAGGAGAGAGCGGGGTACGTGAGAGACCTCTTC 2175
Qy 424 AlaLysThrPheIleValAspGlyLysArgGlyLeuPheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GGCGCGCGCGCTACGTCAGACAGACTGAGAGCGCGGGTG 2214
Qy 444 ArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerPro-----Lys 460
Db 2215 AAGAGCGTGGCGGAGGCGCGGAGCGC---ATGGCTTCAACATGCGCGTCCGGGGCACC 2271

```

```

Qy 461 IleAlaGluMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCCGCCGACATCATGACCTGGCTGATGTGAAGCTCTTCCCGAGGCTGAGGAATGGG 2331
Qy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
Db 2332 GCCAGATGCTCTTCAGCTCCAGCAGCAGCTGCTCTGAGAGCGCCCAAGAGAGGGCG 2391
Qy 498 ThrTyrProAsnGluProGluAlaLysLeuThrPlysArgIleAlaLysIleLeuSer 517
Db 2392 -----GAGCGCTGGCGCGCTGCGCAAGAGATCATGAGAGGGGTATGCC 2439
Qy 518 Leu---LeuProIleGlu 522
Db 2440 CTGGCGCTGCGCGCTGAG 2457

```

RESULT 26  
US-09-350-309-21

Sequence 21, Application US/09350309  
Patent No. 6348314

GENERAL INFORMATION:  
APPLICANT: Prudent, James R.  
Hall, Jeff G.

lyamichay, Victor I.  
TITLE OF INVENTION: Invasive Cleavage of Nucleic Acids

NUMBER OF SEQUENCES: 69  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,309  
FILING DATE: 09-Jul-1999  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/756,386  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 08/599,491  
FILING DATE: 24-JAN-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: FORS-02564  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2502 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 21:  
US-09-350-309-21

Alignment Scores:

Pred. No.: 0.041 Length: 2502  
Score: 106.00 Matches: 119  
Percent Similarity: 33.70% Conservative: 65  
Best Local Similarity: 21.79% Mismatches: 174  
Query Match: 3.91% Indels: 188  
DB: 4 Gaps: 30



us-10-066-551-4 (1-525) x us-09-350-309-21 (1-2502)

QY 118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal----- 130  
 Db 961 GATCTTCTGGCCCTGGCCGCGCCAGGGGGGGGGGTCACCGGGCCCGAGCCTTAT 1020  
 QY 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147  
 Db 1021 AAGCGCTCAGGAGACCTAAGAGAGCGGGGGCTTCGCCAAGACCTGAGCCTTCG 1080  
 QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163  
 Db 1081 GCCCTGAGGAGAGCCTTGGCCTCCCGCCGCGAGACCCCATCTC----- 1128  
 QY 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179  
 Db 1129 -----CTCGCTACCTCTCGAGCCTTCACACACACCCCGAG 1167  
 QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197  
 Db 1168 GGGGTGGCCGCGGCTAGCGGGGGAGTGAAGAGAGAGAGGGGGGAGCGGCCCTT 1227  
 QY 198 -----LeuGlyLysArgAsnIleGlyAspGlu----- 206  
 Db 1228 TCCGAGAGCCTCTCCGACACCTGTGGGGGAGCTTGAAGGGAGAGAGCCTCTTGG 1287  
 QY 207 ---TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223  
 Db 1288 CTTACCGGGAGGTGAGAGAGGCGCTTCCGCTGTGGGCCACATGAG-----GCC 1341  
 QY 224 ThrGly-----SerValAlaGlyLys 230  
 Db 1342 ACGGGGTGGCGGCTGAGGTGAGCTATCTAGAGCCTTGTCCGTGAGGTGGCGGGAG 1401  
 QY 231 ValSer-----HisAspPheAspArgTyrTrp 239  
 Db 1402 ATCGCCGCTCGAGCGGAGTCTTCCGCTGCGCGGCGACCCCTTCAAC----- 1452  
 QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259  
 Db 1453 ---CTCAACTCCCGGAGCAGCTGGAAGAGGTCTCTTGAC-----GAGCTAGGG 1500  
 QY 260 LeuGlnAlaLeuGlyTyrAsnAspGlyThrSerArgHisAla----- 273  
 Db 1501 CTTCGCGCATCGGACAGAGAGAGACCGGACGCTCCACAGCGCCGCTCTCG 1560  
 QY 274 -----LeuLeuArgTyrArgGlyThrVal 281  
 Db 1561 GAGGCGCTCGCGAGGCGCCACCCATCTGAGAGATCTCACTACCGGAGCTCACC 1620  
 QY 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294  
 Db 1621 AAGCTGAAGAGACACCTACATTGACCCCTTCCGAGCTCATCTCACCACCGGCGCGC 1680  
 QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309  
 Db 1681 CTCACACCGCGCTCAACACAGAGCGCACGGGACAGGAGGCTAAGATCCGATCCC 1740  
 QY 310 AlaLysGlyLeuAspArgAspArgLysProIleAlaGlyArgLeuGlnAspAla 329  
 Db 1741 -----AACCTTCAGACATCCCGCTCGCACCCCTTGGGCAAGAGATCCGCGGGCC 1794  
 QY 330 LeuLysGlnProGlyLysSerValTyrLeuValSerProTyr----- 343  
 Db 1795 TTCATCGCGCGAGAGGGGTGCTATGGTGGCCCTGAGCTATAGCCACATAGCTCAGC 1854  
 QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363  
 Db 1855 GTGCTGGCCACACCTCTCGGAGAGAACTGATCGGCTCTCCAGAGAGCGGGGAGC 1914  
 QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377  
 Db 1915 ATCCACAGGAGACCGGACGCTGATGTTCGGCTCCCGGAGGCGGTGAC----- 1968

QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393  
 Db 1969 -----CCCTGATGCGCGGGCGCCAGAGACC 1995  
 QY 394 -----GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406  
 Db 1996 ATCAACTTCGGGGTCTCTAGCGCATGTGGCCACACCGCTCTCCAGAGAGCTACCATTC 2055  
 QY 407 Pro----- 407  
 Db 2056 CCTTACGAGAGAGCCAGGCTTTCATTGAGCCTTTCAGAGCTTCCCAAGGTGGCG 2115  
 QY 408 -----AlaThrLysAspLysGlyLeuThrGlnLysSerValThrSerLeuHis 423  
 Db 2116 GCCTGATTTGAGAGACCTTGAGAGGGGAGAGAGGGGGGTACGTGAGACCTCTTC 2175  
 QY 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443  
 Db 2176 -----GGCGCGCGCTACGTGACAGCTTACAGAGCCTTACAGGCGCGGCTC 2214  
 QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerPro-----Lys 460  
 Db 2215 AAGAGCTGCGCGAGCGCGCCAGCGC---ATGGCTTCAACATGCCCTCCGGGGCACC 2271  
 QY 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477  
 Db 2272 GCGCGGACCTCATAGAGCTGTATGTAAGCTTCTCCAGAGCTGAGGAAATGGGG 2331  
 QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497  
 Db 2332 GCCAGATGCTCTCTAGTCCACAGAGAGTGTCTCTGAGGCCCAAAAGAGGGGG 2391  
 QY 498 ThrTyrProAsnGluProGlnAlaLysLeuTyrLysArgIleAlaLysIleLeuSer 517  
 Db 2392 -----GAGCGCGTGGCCGCGGTGGCAAGAGATCATGAGGGGTATATCCC 2439  
 QY 518 Leu---LeuProIleGly 522  
 Db 2440 CTGGCGGTCGCCCTGAG 2457

RESULT 27  
 US-08-520-946-21  
 ; Sequence 21, Application, US/08520946  
 ; Patent No. 6372424  
 ;  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BROW, MARY ANN D.  
 ; APPLICANT: LYAMICHEV, VICTOR I.  
 ; APPLICANT: OLIVE, DAVID M.  
 ; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF  
 ; TITLE OF INVENTION: PATHOGENS  
 ; NUMBER OF SEQUENCES: 160  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: MEDLEN & CARROLL  
 ; STREET: 220 MONTGOMERY STREET, SUITE 2200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: UNITED STATES OF AMERICA  
 ; ZIP: 94104  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/520,946  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: CARROLL, PETER G.  
 ; REGISTRATION NUMBER: 32,837  
 ; REFERENCE/DOCKET NUMBER: FORS-01756  
 ; TELECOMMUNICATION INFORMATION:



STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States Of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/57,653  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ingolia, Diane E.  
 REGISTRATION NUMBER: 40,027  
 REFERENCE/DOCKET NUMBER: FORS-02565  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (415) 705-8410  
 TELEFAX: (415) 397-8338  
 INFORMATION FOR SEQ ID NO: 187:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2505 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc = "DNA"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..2499  
 US-08-757-653-187

Alignment Scores:  
 Pred. No.: 0.0411 Length: 2505  
 Score: 106.00 Matches: 119  
 Percent Similarity: 33.70% Conservative: 65  
 Best Local Similarity: 21.79% Mismatches: 174  
 Query Match: 3.91% Indels: 188  
 Gaps: 30

US-10-066-551-4 (1-525) x US-08-757-653-187 (1-2505).

QY 118 AsnLeuVal1TYrLeuAlaAlaGluArgGlyValArgVal----- 130  
 Db 961 GATCTTCGGCCCTGCGCCGCCAGGGGGGGCGGTCCACCGGCCCCCGACCTTAT 1020  
 QY 131 ArgLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147  
 Db 1021 AAGCCCTCAGGAGCCTGAGAGGCGGGGGCTTCGCCAAGACCTGAGCTTCTG 1080  
 QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163  
 Db 1081 GCCCTGAGGAGGAGCCCTGGCCCTCCGCCGCCGAGACCCCATGCTC----- 1128  
 QY 164 ValLeuArgLysTrpArgAlaLeuGlyTYrLeuThrAsp-----PheProArg 179  
 Db 1129 -----CTCGCCTACCTCTCGACCCCTTCCACACACACCCCGAG 1167  
 QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197  
 Db 1168 GGGGTGCGCCGCGCTACGCGGGAGAGTGGAGAGGCGGGAGCGGGCGCCCTT 1227  
 QY 198 -----LeuGlyArgAsnIleGlyAspGlu----- 206  
 Db 1228 TCCGAGAGGCTTTCGCCAACCTGTGGGGAGCTTGAAGGGGAGAGGCTCCTTGG 1287  
 QY 207 ---TYrPheValValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223  
 Db 1288 CTTTACCGGAGGTGAGAGGCGCCCTTTCGCGTCTCTGCGCCACATGGAG-----GCC 1341  
 QY 224 ThrGly-----SerValValGlyGlu 230

Db 1342 ACGGGGTGGCGCTGACGTCGCTATCTAGGGCCTTTCCTGAGAGTGGCGGGAG 1401  
 QY 231 ValSer-----HisAspPheAspArgTYrTrp 239  
 Db 1402 ATCGCCCGCTCCGAGCGCGAGGTCTTCGCGTGGCGGCACCCCTTCAAC----- 1452  
 QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGly 259  
 Db 1453 ---CTCACTCCCGGAGACAGCTGGAAGGGTCTCTTTCAC-----GAGCTAGGG 1500  
 QY 260 LeuGlnAlaLeuGlyTYrAsnAspGluThrSerArgHisAla----- 273  
 Db 1501 CTTCGCCGATCGGCAAGAGGAGACCGGACCGCTCCACGAGCGCGCTCCTG 1560  
 QY 274 -----LeuLeuArgTYrArgGlyThrVal 281  
 Db 1561 GAGGCCCTCCGAGAGCCACCCATCTGAGAGAAATCTCGACGTACGCGGAGCTCACC 1620  
 QY 282 GluGln-----SerProLeuTYrGlnLysIle-----GlnThrArg 294  
 Db 1621 AAGCTGAAGACACCTACATTGACCCCTTGGCGGACCTCATCCACCCAGAGGAGCCCG 1680  
 QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309  
 Db 1681 CTCACACCCGCTTCAACAGACGCGCAGCGGCGGAGGAGGCTAGTACGTCGATCC 1740  
 QY 310 AlaLysGlyLeuAspArgAspArgGlyLysProIleAlaGlyArgLeuGlnAspAla 329  
 Db 1741 ---AACCTCCAGAACATCCCGCTCCGACCCGCTTGGCGGAGAGATCCCGCGGCG 1794  
 QY 330 LeuLysGlnProGluLysSerValTYrLeuValSerProTYr----- 343  
 Db 1795 TTCATCCCGGAGAGGAGGGGTGCTATTGTGGCCCTGATAGCCAGATAGACTCAG 1854  
 QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaGlnAspGlyIleAsp 363  
 Db 1855 GTGCTGGCCACCTCTCCGCGGAGAACCTGATCGGGTCTTCCAGAGGGCGGGAG 1914  
 QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377  
 Db 1915 ATCCACAGGAGACCGCAGCTGATGTCGGCTCCCGCGGAGGCGCTGGAC----- 1968  
 QY 378 AlaValHisSerGlyTYrValLysTYrArgLysProLeuLeuLysAla----- 393  
 Db 1969 -----CCCTATGCGCGCGCGCGGCGGCGCAAGACC 1995  
 QY 394 -----GlyIleLysLeuTYrGluLeuGlnProAsnHisAlaVal 406  
 Db 1996 ATCAACTCGGGGTCTTACGCGCATGCGGCCACCGCCTCCAGAGAGTACGATC 2055  
 QY 407 Pro----- 407  
 Db 2056 CCTTACGAGAGGCGGAGCTTATGAGCGCTTTCAGAGCTTCCCAAGTGGCG 2115  
 QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423  
 Db 2116 GCCTGATTTGAGAAGACCTGAGAGGCGGAGGCGGGGTACTGTGAGACCTCTTC 2175  
 QY 424 AlaLysThrPheIleValAspArgLysArgIlePheIleGlySerPheAsnLeuAspPro 443  
 Db 2176 -----GGCGCGCGCGCTACGTGACAGACTAGAGGCCCGGGTG 2214  
 QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerPro-----Lys 460  
 Db 2215 AAGAGCGTGGGAGGCGCGCGAGCGC-----ATGGCTTCAACATGCGCTCCAGGCGACC 2271  
 QY 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTYrAla 477  
 Db 2272 GCGCGGACCTCATGAGCTGATGTGAGACTTTCGCCAGGCTGAGAGAAATGGGG 2331  
 QY 478 TYrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497



```

QY 330 LeuYsglnProGluLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCAATCCGCGAGAGGGGTGGTATATTGGTGGCCCTGAGCATAGACATAGAGCTCAG 1854
QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaValGlnAspGlyLeu 363
Db 1855 GTGCTGGCCACCTCTCCAGGAGAGAACCTGATCCGGTCTTCAGAGAGGGGGAGC 1914
QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACAGGAGAGACCCAGCTGATGTTCCGCCCTCCCGGAGAGCCGTGAGC----- 1968
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGCCGGCGGAGCCCAAGACC 1995
QY 394 -----GlyLeuLysLeuTyrGluLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTCGGGGTCTCTACGGCAATGCGGCCACCCCTCTCCAGAGACTTACCATC 2055
QY 407 Pro----- 407
Db 2056 CCTTACGAGAGAGCCCGCCTTCAATGAGCGCTACTTTCAGAGCTTCCCAAGTCCGG 2115
QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerValThrSerLeuHis 423
Db 2116 GCCTGATTGAGAGACCTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2175
QY 424 AlaLysThrPheLeuAlaValLysArgLysArgLysPheLeuGlnProLysLeuAsp 443
Db 2176 -----GGCCGCGCGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2214
QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerPro-----Lys 460
Db 2215 AAGACGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2271
QY 461 IleAlaGlnIleMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCCCGCGCGCATCATGATGAGTGTGTAAGCTCTTCCAGAGCTGAGGAGGAATGGGG 2331
QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
Db 2332 GCCAGGATGCTCTTCAGGTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2391
QY 498 ThrTyrProAsnGluProGluAlaLysLeuTyrLysArgLysAlaAlaLysIleLeuSer 517
Db 2392 -----GAGGCGCTGGCCCGGCTGCCCAAGAGGTATGAGAGGGGTGATGCC 2439
QY 518 Leu---LeuProIleGlu 522
Db 2440 CTGGCGCTGCCCTGAG 2457

```

```

RESULT 30
US-08-759-038-129
Sequence 129, Application US/08/759038
Patent No. 6090543
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamlichev, Victor I.
APPLICANT: Brow, Mary Ann D.
APPLICANT: Dahlberg, James E.
TITLE OF INVENTION: Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,038
FILING DATE: 02-DEC-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 29-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02574
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 397-8338
FAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 129:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2499
US-08-759-038-129
US-08-759-038-129 (1-525) x US-08-759-038-129 (1-2505)
Alignment Scores:
Pred. No.: 0.0411 Length: 2505
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
Gaps: 30

```













DB: 2 Gaps: 27

US-10-066-551-4 (1-525) x US-08-757-653-167 (1-2505)

QY 118 AsnLeuValTyrLeuAlaIleArgLysValAlaPro-----130

DB 961 GATCTTCTGGCCCTGGCCGCGCCGAGGGGGGCGGCTCCACCGGCCCCGAGCCTTAT 1020

QY 131 ArgLeuLeuAspAspAsnThrArgLysLeu-----AspAspLeuLeu 147

DB 1021 AAGCCCTCAGAGGACCTGAGAGGCGCGGGGCTCTCCGCAAGACCTGAGCGTTCTG 1080

QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163

DB 1081 GCCCTGAGGAGAGCCTTGGCTCCGCGCGGAGACACCCCATGCTC-----1128

QY 164 ValLeuArgLysTyrPheAlaLeuGlyTyrLeuThrAsp-----PheProArg 179

DB 1129 -----CTCGCTACCTCTCGGACCTTCCACACCCACCCCGAG 1167

QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197

DB 1168 GGGGTGGCCCGGCGCTTCCAGCGGGGAGTGGACGAGAGGCGGGGAGCGGCCCTT 1227

QY 198 -----LeuGlyGlyArgAsnIleGlyAspGlu-----206

DB 1228 TCCGAGAGGCTCTTCCGCAACCTTGGGGGAGGCTTGGAGGAGAGGAGGCTTGG 1287

QY 207 ---TyrPheLysValIleGlyLys-----AspThrValPhe 216

DB 1288 CTTTACCGGGGAGTGGAGAGGCGCTTCCGCTCTGCGCCACATGAGGCGAGCGGG 1347

QY 217 AlaAspAspIleLeuAlaIleThrGlySerValValIleGluValSerHisAspPheAsp 236

DB 1348 GTGGCGCTGGACGCTGGCTTACTGAGGCGCTTCTCCGAGAGTGGCGGAGAGATGCC 1407

QY 237 ArgTyrTyrPheLys-----HisSerAlaHis 245

DB 1408 CGCCTCAGAGCGAGTCTTCCGCGCGCGGCGCCACCTTCAACCTCACTCCCGGAG 1467

QY 246 AsnAlaThrArgIleIleArgSerCysAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyr 265

DB 1468 CAGCTGAAAGGCTCTTTGAC-----GAGCTAGGGCTTCCGCGCATGCGCAAG 1518

QY 266 AsnAspGluThrSerArgHisAla-----273

DB 1519 ACGGAGAAGACCGGCAAGCGCTCCACAGCGCGCGCTGAGGCGCTCCGAGAGCC 1578

QY 274 -----LeuLeuArgTyrArgLysGluThrValIleGluGln-----283

DB 1579 CACCCCATCGTGAGAGATCTCTGACAGTACCGGAGCTCACCAAGCTGAAAGACACTAC 1638

QY 284 ---SerProLeuTyrGlnLysIle-----GlnThrGlyArgIleAspTyrGlnSerVal 300

DB 1639 ATTGACCCCTTGGCGGACCTCATCCACCCGAGGAGGCGCGCTCCACACCGCTTCAAC 1698

QY 301 GlnThr-----ArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315

DB 1699 CAGACGGCCAGCGGCGGAGGCTAGTACCTCCATGCC-----AACCTCCAGAAC 1752

QY 316 AspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLys 335

DB 1753 ATCCCGCTCCGACCGCTTGGCGAGAGATCCGCGGCGCTTCATCGCGAGAGAGGG 1812

QY 336 SerValTyrLeuValSerProTyr-----PheValProThrLysSer 349

DB 1813 TGGCTATTGGTGGCCCTGAGCTATGACCATAGACTCAGGCTGCGCCACCTCTCC 1872

QY 350 GlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsn 369

DB 1873 GCGGACGAGAACCTGATCCGGGCTCTCCAGAGGCGGCGGACATCCACAGGAGAGCC 1932

QY 370 Ser-----LeuGlnAlaThrAspValAlaAlaValHisSerCysLys 383

DB 1933 AGCTGATGTTCGGCGTCCCGCGGAGGCGCTGAC-----1968

QY 384 ValLysTyrArgLysProLeuLeuLysAla-----393

DB 1969 -----CCCTGATGCGCGCGGCGGCGCAAGACCATCACTCGGGTCTC 2013

QY 394 ---GlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaValPro-----407

DB 2014 TAGCGCATGTGGCGCCACCGCTCTCCAGAGCTAGCATCTTACGAGAGGCCAG 2073

QY 408 -----AlaThr 409

DB 2074 GCCTTATTGAGCGCTACTTTCAGAGCTTCCCAAGTGGCGGCTGATGAGAACCC 2133

QY 410 LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleVal 429

DB 2134 CTGAGAGGAGGCGAGAGCGGGGGTACGTGAGACCTCTTC-----2175

QY 430 AspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsn 449

DB 2176 ---GGCGCGCGCGCTTACGCTCAGACCTAGAGCGCGCGGAGAGCGCTCGGAGGCG 2232

QY 450 ThrGluMetGlyValValIleGluSerPro-----LysIleAlaGluGlnMetGlu 466

DB 2233 GCGGAGCGC---ATGGCTTCAACATGCGCGTCCAGAGGCGCGCGCGCATGAGAG 2289

QY 467 ArgThrLeuAlaAspThrSerPro-----GluTyrAlaTyrArgValThrLeuAsp 483

DB 2290 CTGCTATGTGTAAGCTCTTCCAGCGCTGAGGAGAAATGGGGCCAGATCTCTTCAG 2349

QY 484 ArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLysThrTyrProAsnGluPro 503

DB 2350 GTCACAAAGACGCTGTCTCGAGGCGCCCAAGAGAGAGGCG-----GAGGCC 2397

QY 504 GluAlaLysLeuTyrPheArgIleAlaAlaLysIleLeuSerLeu---LeuProIleGlu 522

DB 2398 GTGGCGCGCGTGGCAAGAGAGTATGATGAGGGGTGATCCCTGCGCGCTGAG 2457

RESULT 36

US-08-757-653-189

Sequence 189, Application US/08757653

Patent No. 5843669

GENERAL INFORMATION:

APPLICANT: Kaiser, Michael W.

APPLICANT: Lyamichev, Victor I.

APPLICANT: Lyamichev, Natasha

TITLE OF INVENTION: Cleavage of Nucleic Acid Using

TITLE OF INVENTION: Thermostable FEN-1 Endonucleases

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medien & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/757,653

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: F0RS-02565

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338







```

QY 274 -----LeuLeuArgTyrArgGluThrVal 281
Db 1561 GAGGCCCTCCGGAGGCCACCCATCGTGGAGAAATCTGTGAGACGGGGAGCTCACC 1620
QY 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294
Db 1621 AAGCTGAAAGAGACCTACATTGACCTTGGCGGAGACCTACACCCCGAGAGGCGGCC 1680
QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspPro 309
Db 1681 GTCCACACCGCTTCAACAGAGGCCACGCCAGCGGAGCTAAGTACCTGATGCC 1740
QY 310 AlaLysGlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCCAGAACATCCCGCTGCCACCGCTTGGGAGAGATCCGCGGGCC 1794
QY 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCATCCGCGAGGAGGGGTGCTATTGCTGGCCCTGTATAGCCAGATAGAGCTCAGG 1854
QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363
Db 1855 GTGCTGCGCCACCTCTCCGGCGAGCAACCTGATCCGGGTCTTCCAGAGGGGGCGGAC 1914
QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACACGAGAGACCGCCAGCTGATGTCGGGGTCCCGCGGAGCCGCTGAGC----- 1968
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuValAla----- 393
Db 1969 -----CCCGTATGGCGCGGGCGCGGCAAGACC 1995
QY 394 -----GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal 406
Db 1996 ATCACTTGGGGGTCTTACGGCATGTGCGCCACCGCTCTCCAGAGAGCTAGCTATC 2055
QY 407 Pro----- 407
Db 2056 CCTTACGAGAGAGCCAGCCTTATGAGCGCTACTTACAGCTTCCCAAGGTGCGG 2115
QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCCTGATGAGAGAGACCTGAGAGGAGGAGGCGGGGTACTGTGAGAGACCTCTTC 2175
QY 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GGCCCGCCGCTACGTGCCAGACTAGAGCGCGGCTG 2214
QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerPro-----Lys 460
Db 2215 AAGAGCGTGGGAGCGCGCGAGCGC---ATGGCTTCAACATGCCCGTCCAGGCGCAC 2271
QY 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCGGCGGACCTCATGAGGTGCTATGTAAGCTTCCCGAGGCTGAGAAATAGGGG 2331
QY 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497
Db 2332 GCCAGATCTCTCTTACGTCCAGCAGAGAGCTGTCTCGAGGCGCCCAAGAGAGGCG 2391
QY 498 ThrTyrProAsnGlnProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSer 517
Db 2392 -----GAGCGCTGGCGCGCGTGGCCAAAGAGATCATGAGAGGGGTATGCC 2439
QY 518 Leu-----LeuProIleGlu 522
Db 2440 CTGGCGGTGCTCCCTGGAG 2457

```

```

RESULT 39
US-08-759-038-106
; Sequence 106 Application US/08759038
; Patent No. 6090343
; GENERAL INFORMATION:

```

```

APPLICANT: Prudent, James R.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamlicher, Victor I.
APPLICANT: Biew, Mary Ann D.
APPLICANT: Dahlberg, James E.
TITLE OF INVENTION: Cleavage of Nucleic Acids
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESS: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,038
FILING DATE: 02-DEC-1996
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/
FILING DATE: 29-NOV-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02574
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2499
US-08-759-038-106
Alignment Scores:
Pred. No.: 0.054
Score: 105.00
Percent Similarity: 33.52%
Best Local Similarity: 21.48%
Query Match: 3.87%
Matches: 2505
Conservative: 116
Mismatches: 183
Indels: 176
Gaps: 27
US-10-066-551-4 (1-525) x US-08-759-038-106 (1-2505)
QY 118 AsnLeuValTyrLeuAlaIleAlaArgGlyValArgVal----- 130
Db 961 GATCTTCTGGCGCTGTGCGCGCGCAGAGGGGGCGGGGTCCACCGGGCCCGGAGCTTAT 1020
QY 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeuLeu 147
Db 1021 AAAGCCCTCAGGAGACCTGAAGAGCGCGGGGCTTCTGCGCAAGACCTGAGAGCTTCTG 1080
QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTAGGAGAGGCTTGGCTCCGCGCGCGAGACCGCCATGCTC----- 1128

```





MOLECULE TYPE: other nucleic acid  
 DESCRIPTION: /desc - "DNA"  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..2499  
 US-08-759-038-131

Alignment Scores:  
 Pred. No.: 0.054 Length: 2505  
 Score: 105.00 Matches: 119  
 Percent Similarity: 33.708 Conservative: 65  
 Best Local Similarity: 21.79% Mismatches: 174  
 Query Match: 3.87% Indels: 188  
 Gaps: 30

US-10-066-551-4 (1-525) x US-08-759-038-131 (1-2505)

```

OY 118 AsnLeuVal1TyrLeuAla1Arg1Val1ArgVal1----- 130
Db 961 GATCTTCTGGCCCTGCGCCGCCAGGGGGGCGGCTCCACCGGCCCGAGCCTTAT 1020
OY 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAAGCCCTCAGGACCTGAGAGGCGGGGGCTTCTCCGCAAGACCTGAGCGTTCTG 1080
OY 148 AlaLeuAspSer-----HisProAsn1Ile1Val1Arg1LeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGGAGGCGCTTGGCTCCGCCCGCGAGACGCCATGCTC----- 1128
OY 164 ValLeuArg1TyrArg1AlaLeuGly1TyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTCGCTACCTCTGACCCCTTCCACACCCACCCCGAG 1167
OY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArg1AlaThrIle 197
Db 1168 GGGGTGGCCCGCGCTACGCGCGGAGTGGACGAGAGAGCGGGGAGGCGGCCCTT 1227
OY 198 -----LeuGlyArgAsn1Ile1AspGlu----- 206
Db 1228 TCCGAGAGGCTCTTCCCAACCTGTGGGGAGGCTTGAAGGGGAGAGAGGCTCTTGG 1287
OY 207 ---TyrPheLysVal1Gly1Luu-----AspThrValPheAlaAspLeuAsp1IleAla 223
Db 1288 CTTTACCGGGAGGTGAGAGGCGCCCTTCCGCTGCTGGGCCACATGAG-----GCC 1341
OY 224 ThrGly-----SerValVal1Gly1Luu 230
Db 1342 ACGGGGGTGGCGCTGAGCTGCTATCTCAGGGCCCTTCTCCGTGAGGTGGCGGGAG 1401
OY 231 ValSer-----HisAspPheAspArg1TyrTrp 239
Db 1402 ATCGCCCGCTGAGCGGAGGTCTTCCGCTGCGCGCGGACCCCTTCAAC----- 1452
OY 240 AlaSerHisSerAlaHisAsnAlaThrArg1Ile1LeuSerGlyAsn1IleGlyGly 259
Db 1453 ---CTCACTCTCCGCGGACCACTGGAAGGCTCTCTTGTGAC-----GAGCTAGGG 1500
OY 260 LeuGlnAlaLeuGly1TyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCCCGCATGGCAGAGGAGAGACCGGCAAGCCCTCACACCGCCCGCGCTCTG 1560
OY 274 -----LeuLeuArg1TyrArg1LuuThrVal 281
Db 1561 GAGGCCCTCCGAGAGGCCACCCCATCTGAGAGATCTCGAGTACCGGGAGCTCACC 1620
OY 282 GluGln-----SerProLeu1TyrGlnLysIle-----GlnThrGlyArg 294
Db 1621 AAGCTGAAGAGACCTACATTGACCCCTTCCGAGCTCATCCACCCCGAGAGGGCGCG 1680
OY 295 IleAspTyrGlnSerValGlnThr-----ArgLeu1LeuSerAspPro 309
Db 1681 CTCACACCCCGCTTCAACCAAGACGGCCACGCGGAGGCTAAGTACGTCCTCCATCCC 1740

```

```

OY 310 AlaLysGlyLeuAspArgAspArgArgLysProPro1IleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCAGAACATCCCGTCCGACCCCGCTTGGGAGAGGATCCCGCGGCC 1794
OY 330 LeuLysGlnProGluLysSerVal1TyrLeuVal1SerProTyr----- 343
Db 1795 TTCATGCGCGAGAGGGGTGGCTATGTTGGCGCTGCTGATAGCCAGATAGCTCAGG 1854
OY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGly1IleAsp 363
Db 1855 GTGCTGGCCCACTCTCGCGGAGAGAACCTGATCCGGGCTCTCCAGGAGGGCGCGGAC 1914
OY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspVal1Ala 377
Db 1915 ATCCACAGGAGACCGCCAGCTGATGTCGGGCTCCCGGAGAGGCCGTGAGC----- 1968
OY 378 AlaValHisSerGly1TyrValLys1TyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGCGCGCGCGGCCCAAGACC 1995
OY 394 -----Gly1IleLysLeu1TyrGlnLeuGlnProAsnHisAlaVal 406
Db 1996 ATCACTTGGGGTCTCTTACGGCATGCGGCCACCGCTCTCCAGAGACTAGGCTATC 2055
OY 407 Pro----- 407
Db 2056 CCTTACGAGAGGCGCGGCTTATTGAGCGCTTACTTACAGCTTCCCAAGTGGCGG 2115
OY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerVal1ThrSerLeuHis 423
Db 2116 GCCTGATTGAGAAAGCCCTGGAGGAGGAGGAGGCGGGGCTGACCTGGAGACCTCTTC 2175
OY 424 AlaLysThrPheLeuValAspGlyLysArg1IlePhe1IleGlySerPheAsnLeuAspPro 443
Db 2176 -----GCGCGCGCGCTTACGCTGACAGCTTACAGCCCTAGAGCCCGGGTG 2214
OY 444 ArgSerAlaArgLeuAsnThrGluMetGlyVal1IleGluSerPro-----Lys 460
Db 2215 AAGAGCTCGCGGAGCGCGCGGAGCC---ATGGCTTCAACATGCCGCTCCAGGGCACCC 2271
OY 461 IleAlaGlnMetGluArg1ThrLeuAlaAspThrSerPro-----GluTyrAla 477
Db 2272 GCGCGGACCTCATGAAGCTGTGATGTGAACCTTCCCGAGGCTGAGGAAATGGGG 2331
OY 478 TyrArgVal1ThrLeuAspArg1HisAsnArgLeuGln1ThrHisAspProAlaThrArgLys 497
Db 2332 GCCAGATGCTCTTCAAGTCCACGACGAGCTGTCTCGAGGCCCCCAAAAGAGAGGCG 2391
OY 498 ThrTyrProAsnGluProGluAlaLysLeuTyrLysArg1IleAla1AlaLys1IleLeuSer 517
Db 2392 -----GAGCGCGTGGCGCGGCTGCCCAAGAGAGTATGAGGGGGGTGTATCCC 2439
OY 518 Leu-----LeuPro1IleGlu 522
Db 2440 CTGGCGGTGCGCCCTGAG 2457

```

Search completed: May 19, 2003, 18:20:09  
 Job time : 207 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 18:20:19 / Search time 144 Seconds  
(without alignments)  
4702.900 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 2713

Sequence: 1 MRANPKQAMPSETISLMKT.....KLMRIAKILSLPIESL 525

Scoring table: BIOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV=xlh  
-MODE=frame+ p2n model -DB=Published\_Applications\_NA -OFMT=fastap -SUFFIX=trmpb -MINMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Olosum62  
-TRANS=Numan40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US10066551.ecgn.1.1.96.athnat.12052003.091145.22450  
-NCPU=6 -ICPU=3 -NO\_XIPXY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCIT\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/PCITUS\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	628	23.1	1630	US-09-881-752A-355	Sequence 355, App
2	306.5	11.3	4403	US-08-781-986A-78	Sequence 78, Appl
3	280.5	10.3	1446	US-09-974-300-1078	Sequence 1078, Ap
4	275	10.1	1482	US-09-828-523A-21	Sequence 21, Appl

5	275	10.1	1482	US-09-966-521-17	Sequence 17, Appl
6	275	10.1	1509	US-09-828-523A-89	Sequence 89, Appl
7	275	10.1	1509	US-09-966-521-83	Sequence 83, Appl
8	248.5	9.2	28690	US-09-070-927A-138	Sequence 138, App
9	243.5	9.0	1500	US-09-738-626-3014	Sequence 3014, Ap
10	243.5	9.0	1850	US-09-848-726-1	Sequence 1, Appl
11	235	8.7	1183	US-08-781-986A-476	Sequence 476, App
12	208	7.7	640681	US-09-974-300-1093	Sequence 1093, Ap
13	197.5	7.3	933	US-09-974-300-1093	Sequence 415, App
14	162	6.0	6285	US-09-974-300-5529	Sequence 5529, Ap
15	161	5.9	738	US-09-974-300-5529	Sequence 2104, Ap
16	118	4.3	3972	US-09-938-842A-2104	Sequence 545, App
17	113	4.2	1149	US-09-841-132-545	Sequence 2045, Ap
18	112	4.1	3612	US-09-938-842A-2045	Sequence 1375, Ap
19	109.5	4.0	465	US-09-974-300-5542	Sequence 5542, Ap
20	109	4.0	2502	US-10-033-297-21	Sequence 21, Appl
21	106	3.9	2502	US-10-081-806-21	Sequence 21, Appl
22	106	3.9	2502	US-10-074-328-21	Sequence 21, Appl
23	106	3.9	2502	US-09-940-244-21	Sequence 21, Appl
24	106	3.9	2502	US-09-940-244-21	Sequence 21, Appl
25	106	3.9	2502	US-09-940-244-21	Sequence 21, Appl
26	106	3.9	2505	US-10-033-297-68	Sequence 68, Appl
27	106	3.9	2505	US-09-940-244-68	Sequence 68, Appl
28	106	3.9	11739	US-09-070-927A-150	Sequence 150, App
29	105	3.9	2505	US-10-033-297-65	Sequence 65, Appl
30	105	3.9	2505	US-10-033-297-70	Sequence 70, Appl
31	105	3.9	2505	US-09-940-244-65	Sequence 65, Appl
32	105	3.9	2505	US-09-940-244-70	Sequence 70, Appl
33	103.5	3.8	1461	US-09-815-242-9187	Sequence 9187, Ap
34	103	3.8	1377	US-09-938-842A-989	Sequence 989, App
35	103	3.8	2693	US-09-883-825-50	Sequence 50, Appl
36	102.5	3.8	1518	US-10-010-943-3	Sequence 3, Appl
37	102.5	3.8	1917	US-10-010-943-3	Sequence 1, Appl
38	102.5	3.8	1948	US-09-813-153-71	Sequence 71, Appl
39	102.5	3.8	1964	US-10-245-103-41	Sequence 41, Appl
40	102.5	3.8	1964	US-10-245-103-41	Sequence 41, Appl
41	102.5	3.8	1964	US-10-245-143-41	Sequence 41, Appl
42	102.5	3.8	1964	US-10-245-143-41	Sequence 41, Appl
43	102.5	3.8	1964	US-10-245-851-41	Sequence 41, Appl
44	102.5	3.8	1964	US-10-245-851-41	Sequence 41, Appl
45	102.5	3.8	1964	US-10-237-535-41	Sequence 41, Appl

#### ALIGNMENTS

RESULT 1  
US-09-881-752A-355  
Sequence 355, Application US/09881752A  
Patent No. US20020115078A1  
GENERAL INFORMATION:  
APPLICANT: Kleantous, Harold  
APPLICANT: Al-Garawi, Amal  
APPLICANT: Miller, Jean-Francois  
APPLICANT: Tomb, Jean-Francois  
TITLE OF INVENTION: Identification of Polynucleotides  
TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in t  
FILE REFERENCE: 06132/041002  
CURRENT APPLICATION NUMBER: US/09/881,752A  
CURRENT FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 08/833,457  
PRIOR FILING DATE: 1997-04-01  
NUMBER OF SEQ ID NOS: 370  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 355  
LENGTH: 1630  
TYPE: DNA  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (294)...(1577)  
US-09-881-752A-355

## Alignment Scores:

Pred. No.: 3,16e-64 Length: 1630  
 Score: 628.00 Matches: 163  
 Percent Similarity: 48.85% Conservative: 93  
 Best Local Similarity: 31.11% Mismatch: 170  
 Query Match: 23.15% Indels: 98  
 Gaps: 19

US-10-066-551-4 (1-525) x US-09-881-752A-355 (1-1630)

QY 52 ThreSerlyspProValArgLeuaspasnIleuGlnIleArgHISThProHISThrAsn 71  
 DB 147 TCAAGCTCTCTATCTTATGAT-----CCCTACACTACC 182  
 QY 72 GlySerAspIleTyr----- 77  
 DB 183 CCCATTTGGAGCTTGTATGCTGGAATAAATTAAGAAACCCCTAACCATAGCGGCCATT 242  
 QY 78 LeuLeuaspProHISGlnAlaIleAlaIleArgAlaIleuIleGluSerAlaGlu 97  
 DB 243 CTTTATGAAAGTGGCTTGTACGCTCTGTGCATAGAGTGGCTTATTAAGATGAGCCAA 302  
 QY 98 HIsSerLeuaspIleuGlnTyrTyrIleTrrparGaspaspIleSerGlyArgLeuIlePhe 117  
 DB 303 AAAAGCATGACATGCAACTTATATCTATATAAAGACGCTTCTTCTCAAGTGTGCT 362  
 QY 118 AsnLeuValTyrLeuAlaIleAlaGlnArgGlyValArgValArgLeuLeuaspasn 137  
 DB 363 AAAGACCTTTAAATGCGCCCAATCGTGGGTAAAGTCCGATCCTTTAGACATATAC 422  
 QY 138 AsnThrArgGlyLeuasp-----AspLeuLeuAlaIleuaspSerHISPro 153  
 DB 423 -----GGATTGGATTGCGATTTCAGATATATATGCTC---TTAAATTTCCATAAA 470  
 QY 154 AsnIleGluValArgLeuPheaspProPheValLeuArgLysTrrparAlaLeuGlyTyr 173  
 DB 471 AACATTGAGTGAATAATTTTAAACCTTACTATATACCGCAAT---AAAGCCTTGGCTTAT 527  
 QY 174 -----LeuThrAspPheProArgLeuaspasnArgArgMetHISAsnLysSerPheThr 190  
 DB 528 TTTGAATGCTTGGCGATTATGAGCGCATTAATAAACCCATGACACACAGCTTTTATC 587  
 QY 191 AlaaspasnArgAlaIleHISLeuGlyValArgasnIleGlyaspGlyTyrPheLysVal 210  
 DB 588 GTGATATATTTGCTGTCATATAGGGGGCGCATATTTGGGACAAATATTGATTAAC 647  
 QY 211 GlnGluaspThrValPheAlaaspLeuaspIleLeuAlaIleHISLysSerValValGlyL 230  
 DB 648 GATTTAGACACGAAATTTTATGATTAGACGCTTTGTTGGGGGGGCTTCAAAA 707  
 QY 231 ValSerHISaspPheaspArgTyrTrrparLysSerHISAsnAlaHISAsnAlaThrArgL 250  
 DB 708 GCCAAAGAAAGCTTGAACGCTATGAGATTCACCGCTCATCCGTTTCA---TTA 764  
 QY 251 IleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnaspGlnThrSer 270  
 DB 765 CTAAGAAC----- 773  
 QY 271 ArgHISAlaLeuLeuArgTyrArgGluThrValGlnGlnSerProLeuTyrGlnLysIle 290  
 DB 774 ---CATAAAGACTCAAA---AACACGCTTAAGAAATCGCTTAACCTCATGAATAAAATC 827  
 QY 291 GlnThrGlyArgIleAspTrrparGlnSerValGlnThrArgLeu----- 304  
 DB 828 CCTATCAGCGCTGAAGACAAACACAGTTTGAATAAAGCATGATTATATAGATCGT 887  
 QY 305 -----LISerAspaspPro 309  
 DB 888 TTCCAAATAATCAATACCCATTATATGAGATGCAATTTTATGCGGATTCACCC 947  
 QY 310 AlaLysGlyLeuaspArgaspArgArgLysProPheIleAlaGlyArgLeuGlnAspAla 329

DB 948 AAAAAA---ATTGACACGCCCTTGATTGCG---CCTATCAAAATCGCTTTTGAGAAAGCC 1001  
 QY 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLysSer 349  
 DB 1002 CTTTAAAGACGCTAAGAGACTCGTTTATCGCTTGTATGATTTATTCACAGGCAAAAG 1061  
 QY 350 GlnThrAspAlaLeuAlaLysLeuValGlnaspGlyIleaspValThrValLeuThrAsn 369  
 DB 1062 ATGATGAATAATCTTAAATCAAAATTTCTAAGGAGGATGTAATGACATCTTACCAAT 1121  
 QY 370 SerLeuGlnAlaIleThrAspValAlaAlaValHISSerGlyTyrValLysTyrArgLysPro 389  
 DB 1122 TCCCTTATCATCTACGATGAGCATAGTACTGCTATAGGGGATGGGAAAGTATCGCAACCA 1181  
 QY 390 LeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHISAlaValProAlaThr 409  
 DB 1182 TTAGTGCATAGTGGCGCAATGTCTATAAATACGAACGATTTTTCACCGCCAGATT 1241  
 QY 410 LysaspLysGlyLeuThrGlySerSerValThrSerHISAlaLysThrPheIleVal 429  
 DB 1242 AAAGGCGCC-----TTTAGCACCAAAACATTCCTGCATGGCAAGACGATTTGTTT 1292  
 QY 430 AspGlyLysArgIlePheIleGlySerPheAsnLeuaspProArgSerAlaArgLeuasn 449  
 DB 1293 GATGACAAATTAACGCTTCTAGGGAGTTTCATATGATTCGCGCTGTACATACATCAC 1352  
 QY 450 ThrGluMetGlyValAlaIleGluSerProLysIleAlaGlnIleMetGluArgThrLeu 469  
 DB 1353 ACTGAACCGCGGTTTGTGTTGACACACCGCTTTGTGTAAGAAGCGCTTGTGCGCTT 1412  
 QY 470 AlaaspThrSerProGluTyrAlaTyrValThrLeuaspArgHISAsnArgLeuGln 489  
 DB 1413 AAAGATCATGCC---CAACAATCATAGCATTTGGGTGTATCGCAT---AGAGTGATT 1466  
 QY 490 Trp-----HISaspProAlaThrArgLysThrTyrPro 500  
 DB 1467 TGGAGACGGGTGGAAGAACGATTTTAATCCATGAA-----AAACT----- 1508  
 QY 501 AsnGluProGluAlaLysLeuTrrparLysArgIleAlaLysIleLeuSerLeuPro 520  
 DB 1509 ---TCGCCTGACACTTCTCTTTTGGCGTTGATTAAAGAATGCTTAAGTCTTCTCT 1565  
 QY 521 IleGluSerLeu 524  
 DB 1566 GAAAGACAGCTT 1577

RESULT 2  
 US-08-781-986A-78  
 Sequence 78, Application US/08781986A  
 Publication No. US2003005436A1  
 GENERAL INFORMATION:  
 APPLICANT: Charles Kunsch  
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
 NUMBER OF SEQUENCES: 5255  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/781,986A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:



```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1078
: LENGTH: 1446
: TYPE: DNA
: ORGANISM: Bacillus licheniformis
US-09-974-300-1078

```

Alignment Scores:	
Pred. No.:	4, 13e-23
Score:	1446
Percent Similarity:	280.50
Best Local Similarity:	39.59%
Query Match:	24.89%
DB:	10.348
Length:	1446
Matches:	110
Conservative:	65
Mismatches:	152
Indels:	16
Gaps:	11

OY	76	ILTYLLEULEAASAPPROHISGLIALPHELIAlALatrgalialaleuileguler	95
Dd	360	GTCGAGCGCTTACGGACGAGCACCAGAATTCGACAGCGCTGCCGATATTGAANAAC	419
OY	96	AlActunHisSerLeuAspPheGlnTYrTYrILEttrArgAsnSRIeserGIyArgLeu	115
Dd	420	GCCAAAGACCATATCATCTTCATATTTATTTATTAAGCGCATGTAGATCGCAANAAG	479
OY	116	LeuPheAsnLeuValTYrLeuAlAlaIGluArgGlyIyAlArGyAlArgLeuLeuAsp	135
Dd	480	CTCCGGGATGGCGTCATCAAAAAAGCAAAAAMAGCGCTGAAGTCCCGTTTTGTATGAT	539
OY	136	AspAsnAnThrArgGlyIyLeuAspPheLeuLeuAlaLeuAspSerHisProAsnIle	155
Dd	540	GAATTAGGATCAAGAGCGCTCAG-----	563
OY	156	GIUValArgLeuPheAsnProPheValLeuArgLYstrArgAlaLeuGIyTYrLeu---	174
Dd	564	-----AAAAGCTTTTCAAAGACTGAAGCTGA AAAAGCGGAGGATTTTGTCGAA	608
OY	175	ThraSPhePro-----ArgLeuAsnArgMetHisAsn	186
Dd	609	GTCCTTTTCCGCGCGCTTAATTTATTCATTTGGCGCATGATTCGCCAACACCGGC	668
OY	187	LysSerPheThrAlaAspAsnArgAlaThrILEuGIyGlyArgAsnILEgIyAspGI	206
Dd	669	AAGCTCGTCATTATCGACGGCGTCATCGCATGTGCGAGGATTTAATGTGCGAGACGAA	728
OY	207	TyrPheLySValIGluAspThrValPheLa-----AspLeuAsPIleLeuAla	223
Dd	729	TAT-----CTCGGCTGAAACCCGCCCTTTGGCTATTGGCGGGCACGCACTTAGATC	782
OY	224	ThrGlySerValValIGluValuIsenHisAspPheAspArgTYrTrp-----Alaser	241
Dd	783	AAAGAACACGGGSCACTGCATTCAMAACGGGTTTTATCGTGATGGATTCAGGCTCG	842
OY	242	HisSerAlaHisAsnAlaThr-----ArgllelleaTyserGly	254
Dd	843	CACCAT---CACGACATTACTATATATGSCAAACCATTTTCTGCATCACTGCCACGGA	899
OY	255	AsnILEgIySGlyLeuGlnAlaLeuGIyTYrLasnaSpGluThrSerArgHisAlaLeu	274
Dd	900	AACATCGGCATGCAAAATGTGTACCACGGCGCCCGCATTCAGAA-----	941
OY	275	LeuArgTYrArgGluThrValIGluGlnSerProLeuTYrGlnIySILEgInThrGIyArg	294
Dd	942	-----TGGGAGCGAGATTAAACCGGTAT	965
OY	295	IleAspTrpGlnInsertValGlnThrArgLeuIeSerAspAspProAlaIySGlyLeuAsp	314
Dd	966	ATT-----AAATGATATCA-----	980
OY	315	ArgspArgArgLysProProIleAlaGIyArgLeuGlInsprAlaLeuIySGlnProGlu	334
Dd	981	-----ACGGCAAG	989

```

Oy 335 LysSerValTYLLeuValSerProTYRPhelValProThylGlySerGlyThrAspAlaLeu 354
Db 990 CGGTCCATCTCTATTACGACGCCCTTATTATTACCGGACGACGACCGCTGCTGATCCCTC 1049
Oy 335 AlaLysLeuValGlnAspGlyIleAspValThrValLeuThraSerLeuGlnAlaThr 374
Db 1050 AGGATCGCGTGTCTGTCTCCGCGTGTGACGTCGAGCATGATTCCTTAAC-----AAACCC 1103
Oy 375 AspValAlaAlaValaHisSerGlyTYrValLysTYrArgLysProLeuLeuLysAlaGly 394
Db 1104 GACCATCCGTTTGTGTACTGGGCAACCTATTCTTACATCGCGGAACCTTTTAAAGACAGCC 1163
Oy 395 IleLysLeuTYrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeu 414
Db 1164 GCCTCCGATTT-----ATTTACGACACAGGTTTT 1193
Oy 415 ThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgTle 434
Db 1194 -----ATTCATGCGAAACGATTTGTAGTGAAGAGAAATTTTCA 1232
Oy 435 PheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyVal 454
Db 1233 TCGGTCGCGACGCGCAAAATTATTGATTATGCGGACGCTTCAACCTCAATTTTGAAGTCA 1292
Oy 455 ValIleGlySerProLysIleAlaGluGlnMet-----GluArgThrLeuAla 470
Db 1293 TTTTATTATGATGAGGAAATTGCGAAGCCTTGCTTCATTGTAAGAAAGACTG--- 1349
Oy 471 AspThrSerProGluTYrAlaTYrAlaTYrValaIThrLeuAspArgHisAsnArgLeuGlnTrp 490
Db 1350 CAAATGTCACAGGAGCGTACGACCTTTGAGGAAATTATTAAACAGAGCAATTCATCGCTTC 1409
Oy 491 HisAsp 492
Db 1410 AAAGAA 1415

RESULT 4
US-09-828-523A-21
; Sequence 21. Application US/09828523A
; Patent No. US20020168697A1
; GENERAL INFORMATION:
; APPLICANT: The Pharmacia & Upjohn Company
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 268. 62120101
; CURRENT APPLICATION NUMBER: US/09/828,523A
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/266,327
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: Patent version 3.1.
; SEQ ID NO 21
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-828-523A-21

Alignment Scores:
Pred. No.: 1,93e-22 Length: 1482
Score: 275.00 Matches: 100
Percent Similarity: 38.848 Conservative: 81
Best Local Similarity: 21.466 Mismatches: 149
Query Match: 10.146 Indels: 136
DB: 9 Gaps: 16

US-10-066-551-4 (1-525) x US-09-828-523A-21 (1-1482)
Oy 74 SerAspIleTYr-----LeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAla 91
Db 415 AATGATTTATATGATCAAGATTTTAAAAAGAT----- 444
Oy 92 LeuIleGlySerAlaGluHisSerLeuAspLeuGlnTYrTYrIleTrpArgAsnAspIle 111
Db 445 ---ATTTAAATGCAAAAGATATATTCATTTAGAGTACTATATACCTTGCGTTAGATGCT 501

```

```

US-10-066-551-4 (1-525) x US-09-828-523A-21 (1-1482)
Oy 74 SerSpIleTyr-----LeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAla 91
    :::::|||||Tyr ::::| |||
Db 415 AATGATTATATGATCAAGTTTAAAGAT----- 444
Oy 92 LeuIleGluSerAlaGluHisSerLeuAspLeuGluTyrTyrIleTyrArgAsnAspIle 111
    :::::|||||Tyr ::::| |||
Db 445 --ATTAAAAATGCAAAAGATATATCATATTTAAAGATATCTACTTGGCTTGATGGATGT 501
    :::::|||||Tyr ::::| |||

```

QY 112 SerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleArgGlyValArgValArg 131  
 Db 502 TTAGGTAAAGAAATTTTACATGCTTTAGAGAAAGAAATTCAGACAGCTCAGAGTAAAG 561  
 QY 132 LeuLeuLeuAspAspAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer 151  
 Db 562 ATATATATATGATGATGCTTGCATCTAA----- 588  
 QY 152 HisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeu 171  
 Db 589 -----AATGTTAAGATGCAAAATTTTGATCATTTT-----AATCGTTA 627  
 QY 172 Gly-----TyrLeuThrAspPhePro-----ArgLeuAsn 181  
 Db 628 GGTGGAGAGTGAAGCATTTTTCCTTCAAAATTCCTTTATGTAATTCAGAAATGAAAT 667  
 QY 182 ArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArg 201  
 Db 668 AATAGAAATCATAGAAAGAAATCATCGTAATCGATGCTCACTAGCTTATCTCGAGCATTT 747  
 QY 202 AsnIleGlyAspGluTyrPheLysValGlyGluAspThrValPheAlaAspLeuAspIle 221  
 Db 748 AACATTGGTGAATATATAGCATTTAGGAAATTAAGATTTGGAGATACGCATTTTA 807  
 QY 222 LeuAlaThrGlySerValValGlyGluValSerHisAspPheAspArgTyrTrpAlaSer 241  
 Db 808 CGTATACAAAGGGATGCGGTTGATGCACCTGCAGTTCGATTTATTTTACAGTGCAGATTG 867  
 QY 242 HisSerAlaHisAsnAlaThrArgIleIle-----ArgSerGly 254  
 Db 868 ---CAAGCCACCGCTCCACAAATTTGAATGATGATTAGTATCTCCATAAAGAACGGA 924  
 QY 255 AsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 274  
 Db 925 CCATTGGGC----- 933  
 QY 275 LeuArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArg 294  
 Db 934 -----AATTCACCAATTCAAATAGCTCAGTGGCCGGCT 969  
 QY 295 IleAspTrpGlnSerValGln-----ThrArgLeuIleSerAspAspProAlaLys 311  
 Db 970 ACTGACTGCGCATCAAAATGTAATACGCTTATACAAAATGATT----- 1011  
 QY 312 GlyLeuAspArgAspArgArgLysProIleAlaGlyArgLeuGlnAspAlaLeuLys 331  
 Db 1012 -----ATG 1014  
 QY 332 GlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThr 351  
 Db 1015 ACTGCAGAAAGAAATCTGTATATTACATCACCATTATTCGCGATTAATTCATATATA 1074  
 QY 352 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeu 371  
 Db 1075 AATGCCATTAAATTCGCTCCTCAATCAGCTGTAGATGATCATTAATGATTCATCATGT 1131  
 QY 372 GlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeu 391  
 Db 1132 ---AAGCCAGATCATTCATTAATATATGCGGACATTTTCAAAAGCCCTGCTATATA 1188  
 QY 392 LysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAsp 411  
 Db 1189 TCAAGTGTGTTAAATTTAT-----ACGTATGAA 1218  
 QY 412 LysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGly 431  
 Db 1219 AATGATTT-----ATACATTCCTAAATGCTTAATGATGAT 1257  
 QY 432 LysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlu 451  
 Db 1258 GAATCGTATCTAGTGGCAGCAAAATGAGCTTATAGAGTTTGAATTAATTTTGA 1317

QY 452 MetGlyValValIleGluSerProLysIleAlaGluGlnMetGluArgThrLeuAlaAsp 471  
 Db 1318 GTAATATGCTTTGTATATGATGAAGAAATCTGTCAAGATTTA----- 1359  
 QY 472 ThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHis 491  
 Db 1360 -----AGGCTGCTTATGAAACATGATATTACAAAAATCAACAACTA----- 1401  
 QY 492 AspProAlaThrArgLysTrpTyrProAsnGluProGluLysLeuTrpLysArgIle 511  
 Db 1402 -----ACCAAGATATATGCAATAGACCGCTGTCTTAATTCAGAAATCG 1452  
 QY 512 AlaAlaLysIleLeuSer 517  
 Db 1453 TTAGCAAAATTTACTTCG 1470  
 RESULT 5  
 US-09-966-521-17  
 ; Sequence 17, Application US/09966521  
 ; Publication No. US20030087321A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TOMICH, Che-Shen  
 ; APPLICANT: QUINN, Cheryl  
 ; APPLICANT: ARVIDSON, Staffan  
 ; APPLICANT: HARRIS, Douglas  
 ; APPLICANT: MOTT, John  
 ; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS  
 ; FILE REFERENCE: 6212.N2  
 ; CURRENT APPLICATION NUMBER: US/09/966,521  
 ; CURRENT FILING DATE: 2001-09-28  
 ; NUMBER OF SEQ ID NOS: 145  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 17  
 ; LENGTH: 1482  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus aureus  
 ; US-09-966-521-17  
 Alignment Scores:  
 pred. No.: 1,93e-22 Length: 1482  
 Score: 275.00 Matches: 100  
 Percent Similarity: 38.84 Conservative: 81  
 Best Local Similarity: 21.46 Mismatches: 149  
 Query Match: 10.14 Indels: 136  
 DB: 9 Gaps: 16  
 US-10-066-551-4 (1-525) x US-09-966-521-17 (1-1482)  
 QY 74 SerAspIleTyr-----LeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAla 91  
 Db 415 AATGATTTATATGATCATCAGTTTAAAGAT----- 444  
 QY 92 LeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTrpIleTrpArgAsnAspIle 111  
 Db 445 ---ATTAAAGATGCAAGAAAGATATATCCATTGTAGACTATCTTCCGCTTATGATGCT 501  
 QY 112 SerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaIleArgGlyValArgValArg 131  
 Db 502 TTAGGTAAAGAAATTTTACATGCTTTAGAGAAAGAAATTCAGACAGCTCAGAGTAAAG 561  
 QY 132 LeuLeuLeuAspAspAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer 151  
 Db 562 ATATATATATGATGATGCTTGCATCTAA----- 588  
 QY 152 HisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeu 171  
 Db 589 -----AATGTTAAGATGCAAAATTTTGATCATTTT-----AATCGTTA 627  
 QY 172 Gly-----TyrLeuThrAspPhePro-----ArgLeuAsn 181  
 Db 628 GGTGGAGAGTGAAGCATTTTTCCTTCAAAATTCCTTTATGTAATTCAGAAATGAAAT 667  
 QY 182 ArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArg 201

```

Db 668 AATAGAAATCATAGAAAATCATCTGTAATCGATGCTCAACTAGTAGTATCGGAGATT 747
OY 202 AsnilegiaspGluTyrPheIysValGlyIuaspThrValPheAlaAspLeuaspIle 221
Db 748 AACATGGTGATGATATCTAGAGATTAGAAAATTAGCATATGGAGACATACCATTTA 807
OY 222 LeuAlaThrGlySerValValGlyIuValSerHisAspPheAspArgTyrTrpAlaSer 241
Db 808 CGTATACAGAGGGATGCGGTGATGATCGACGTGATGATTTATTTAGATGGAATTCG 867
OY 242 HisSerAlaHisAsnAlaThrArgIleIle-----ArgSerGly 254
Db 868 ---CAAGCGCACCGCTCCCAATTTGAAATATGATGTTAAATTTCCCPAAAAAGAACGA 924
OY 255 AsnilegIyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 274
Db 925 CCATTGGGC----- 933
OY 275 LeuArgTyrArgGluThrValGluIuSerProLeuTyrGlnLysIleGlnThrGlyArg 294
Db 934 -----AATTCACCAATTCAAATAGCTGCAAGTGGCCGCT 969
OY 295 IleAspTrpGlnSerValGln-----ThrArgLeuIleSerAspAspProAlaLys 311
Db 970 AGTACTGCGCATCAATATGAATACGCTTATACAAAATGATT----- 1011
OY 312 GlyLeuAspArgAspArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLys 331
Db 1012 -----ATG 1014
OY 332 GlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThr 351
Db 1015 AGTCGAAAGAAATTCGTATATTACATCCCATTTTCATTCGCGAATTCATATATA 1074
OY 352 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeu 371
Db 1075 AATGCCATTAATTCGTCGTAATACAGTGTAGATGTACATTATATGATTTCCATGT--- 1131
OY 372 GlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeu 391
Db 1132 ---AAGCCAGATCATCCATTAGTATATGCGGACATTTTCAATGCCCTGACTTATATA 1188
OY 392 LysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAsp 411
Db 1189 TCAATGTGTGTTAAATTTAT-----ACGATGAA 1218
OY 412 LysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValaspGly 431
Db 1219 AATGATTT-----ATACATTCTAAATGTGCTTAATTTGATGAT 1257
OY 432 LysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlu 451
Db 1258 GAAATCGTATCAGTGGGACAGCAATATGACCTTAGAGATTGGAATTAATTTTGA 1317
OY 452 MetGlyValValIleGluSerProLysIleAlaGluGlnMetGluArgThrLeuAlaAsp 471
Db 1318 GTAATGCCCTTTATATGATGAAATATCTGTGTAAGAATTTA----- 1359
OY 472 ThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPhe 491
Db 1360 -----AGGTCGCTTATGACATGATTTACAAATCAAAACAATA----- 1401
OY 492 AspProAlaThrArgLysThrTyrProAsnGluProGluAlaLysLeuThrLysArgIle 511
Db 1402 -----ACCAAGATCATATATGCCAATAGACCGCTGTCTTAATTTCAAAAGATCG 1452
OY 512 AlaAlaLysIleLeuSer 517
Db 1453 TTAGCAAAATAGTTTCG 1470

```

RESULT 6  
US-09-828-523A-89

```

; Sequence 89, Application US/09828523A
; Patent No. US2002016897A1
; GENERAL INFORMATION:
; APPLICANT: The Pharmacia & Upjohn Company
; TITLE OF INVENTION: ANTIMICROBIAL METHODS AND MATERIALS
; FILE REFERENCE: 268.62120101
; CURRENT APPLICATION NUMBER: US/09/828.523A
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/266,327
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 99
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 89
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence of S. aureus coding region cloned for expr
; US-09-828-523A-89

Alignment Scores:
Pred. No.: 1.98e-22 Length: 1509
Score: 275.00 Matches: 100
Percent Similarity: 38.84% Conservative: 81
Best Local Similarity: 21.46% Mismatches: 149
Query Match: 10.14% Indels: 136
DB: Gaps: 16

US-10-066-551-4 (1-525) x US-09-828-523A-89 (1-1509)
OY 74 SerAspIleTyr-----LeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAla 91
Db 418 AATGATTTATGATGATCAAGTTTAAACAT----- 447
OY 92 LeuIleGluSerAlaGlnHisSerLeuAspLeuGlnTyrTyrIleTrpArgAsnAspIle 111
Db 448 ---ATTAAAGATGCAAAAGATATATCCATTAGAGTACTTACTTGTGCTTTAGATGCT 504
OY 112 SerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArg 131
Db 505 TTAGTAAAGAAATTTTACATGCTTTAGACAGAAATTTGAAACAGCTGAGACTAATA 564
OY 132 LeuLeuLeuAspAspAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer 151
Db 565 ATATTATATGATGATGTGATCTAA----- 591
OY 152 HisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeu 171
Db 592 -----AATGTTAAGATGCAAAATTTGATCTATT-----AAATCGTTA 630
OY 172 Gly-----TyrLeuThrAspPhePro-----ArgLeuAsn 181
Db 631 GGTGAGAAAGTTGAGCATTTTTCCTTCCAAATATACGCTTATTAATTTGAGATGAT 690
OY 182 ArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArg 201
Db 691 AATAGAAATCATAGAAAATCATCTGTAATCGATGCTCAACTAGTAGTATCGGAGATT 750
OY 202 AsnilegiaspGluTyrPheLysValGlyIuaspThrValPheAlaAspLeuaspIle 221
Db 751 AACATGGTGATGATATCTAGAGATTAGAAAATTAGCATATGGAGACATACCATTTA 810
OY 222 LeuAlaThrGlySerValValGlyIuValSerHisAspPheAspArgTyrTrpAlaSer 241
Db 811 CGTATACAGAGGGATGCGGTGATGATCGACGTGATGATTTATTTAGATGGAATTCG 870
OY 242 HisSerAlaHisAsnAlaThrArgIleIle-----ArgSerGly 254
Db 871 ---CAAGCGCACCGCTCCCAATTTGAAATATGATGTTAAATTTCCCPAAAAAGAACGA 927
OY 255 AsnilegIyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 274

```



QY	Db	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	DB:
US-10-066-551-4 (1-525)	x	US-09-966-521-83 (1-1509)				
QY	74	1.98e-22	Length: 1509			
Db	418	275.00	Matches: 100			
QY	92	38.84%	Conservative: 81			
Db	448	21.46%	Mismatches: 149			
QY	112	10.14%	Indels: 136			
Db	504		Gaps: 16			
QY	505					
Db	564					
QY	132					
Db	565					
QY	152					
Db	592					
QY	172					
Db	631					
QY	182					
Db	691					
QY	202					
Db	751					
QY	222					
Db	811					
QY	242					
Db	871					
QY	255					
Db	928					
QY	275					
Db	937					
QY	295					
Db	973					
QY	312					
Db	1015					

DB 1018 AGTCAAGAAATCTGTATTATTCAATCCACCATTTTCATTCGCGAATATCATATATA 1077  
DB 352 ASPALAEUALATLSLEUVALINASPGLYLEASPVALTIRVALLEUThAsnSerLeu 371  
DB 1078 AATGCCATTAAATTCGCTCAATATCAGGTAGTACATTATATATTCATGT--- 1134  
DB 372 GlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeu 391  
DB 1135 ---AAGCCAGATCATCCATTAGTATATGGCGACATTTTCAATGCCCTGACTATATA 1191  
DB 392 LysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAlaThrLysAsp 411  
DB 1192 TCAAGTGGTGTAAATAATTAT---ACGATGAA 1221  
DB 412 LysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspGly 431  
DB 1222 AATGGATT-----ATCATTTCTAAAGTGGCTTAAATGATCAT 1260  
DB 432 LysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlu 451  
DB 1261 GAATCGATCAGTGGCAGACAAATATGAGCTTTAGAACTTTTGAATTAATTTCGA 1320  
DB 452 MetGlyValValIleGlySerProLysIleAlaGluGlnMetGluArgThrLeuAlaAsp 471  
DB 1321 GTAATGCCCTTGTATATGATGAATAATCTTCCTAAGATTTA----- 1362  
DB 472 ThrSerProGluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnThrHis 491  
DB 1363 -----AGGTGGCTTATGACACATCATATTTACAAATCAACAACTA----- 1404  
DB 492 AspProAlaThrArgLysThrTyrProAsnGluProGluAlaLysLeuThrLysArgIle 511  
DB 1405 -----ACCAAGATCATATATGACCAATAGACCGCTGTCTTAATTCAGAAATCG 1455  
DB 512 AlaAlaLysIleLeuSer 517  
DB 1456 TTACCAAAATTAGTTTCG 1473

RESULT 8  
US-09-070-927A-138/c  
Sequence 138, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070, 927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 138:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 28690 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 138:  
US-09-070-927A-138

Alignment Scores:  
Pred. No: 3,1e-17 Length: 28690  
Score: 248.50 Matches: 105  
Percent Similarity: 38.37% Conservative: 60  
Best Local Similarity: 24.42% Mismatches: 184  
Query Match: 9.16% Indels: 81  
DB: Gaps: 14

US-10-066-551-4 (1-525) x US-09-070-927A-138 (1-28690)

DB 40 ProLeuGluGluArgThrGluSerArgHisPheAsnThrSerLysProValArgLeuAsp 59  
DB 28537 CCGACGGAGGAGCAACAACATTTACAGACTTTTTCCTCG----- 28496  
DB 60 AsnIleLeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIleTyrLeuLeu 79  
DB 28495 -----TTAAATAGAGTCCCTCTGACTAG--AAAATGATGTTGAATTTT 28451  
DB 80 AsnAspProHisGluAlaPheAlaAlaArgAlaLeuIleGluSerAlaGluHisSer 99  
DB 28450 ACGGATGTACGAAATAATTCATATCGCTGATGCGGATATATAAAAGCTCACACTCC 28391  
DB 100 LeuAspLeuGlnTyrTyrIleTyrPArgAsnAspIleSerGlyArgLeuLeuPheAsnLeu 119  
DB 28390 ATTCTATATTGAATATTAATGCTTTGTAAACGATCATATCGCACCAAAATTTGAACCTTA 28331  
DB 120 ValTyrLeuAlaIleGluArgGlyValArgValArgLeuLeuAspAspAsnThr 139  
DB 28330 TTACAGCAAAAGCGCGCTGAAAGGGGTGAAGTGGTGTTCGATCGACTTGGCTCA 28271  
DB 140 ArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeu 159  
DB 28270 AAGGCC-----ACCAAGTTCATCATTTTGATGATGAAATTA 28232  
DB 160 PheAsnProPheValLeuArgLysTyrPArgAlaLeuGlyTyrLeuThrAspPheProArg 179  
DB 28231 AACGGTGGTTTGTCCAAACCTTTATTACTCTCAAAAAGCACCTTTGAACTTT---CGT 28175  
DB 180 LeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGly 199  
DB 28174 TTGAATTATCATGATCAACCGGAAAATTGTGTATTTGACGGAAGAGTACACTTGGC 28115  
DB 200 GlyArgAsnIleGlyAspGluTyrPheLysValGlyGluAspThrValPheAlaAspLeu 219  
DB 28114 GGCCTTATATGTGGCATCATAT-----CCCGAAGCATATAAAAGTTTGGC----- 28067  
DB 220 AspIleLeuAlaThrGlySerValAlaGlyGluValSerHisAspPheAspArgTyrTrp 239  
DB 28066 -----TATTGG 28061  
DB 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259  
DB 28060 CGGGAT-----ACACATTTTAGCATGCAAGGCGAGAGCTCATTA 28019  
DB 260 LeuGln-----AlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArg 276  
DB 28018 CTGCAATGCGTTTTTTAAATGATGAGACGTCCTTCCCGGAGAAATCGTGTGGCG 27959

[illegible]

Alignment Scores:			
Pred. No.:	1,07e-18	Length:	1500
Score:	243.50	Matches:	105
Percent Similarity:	36.24%	Conservative:	61
Best Local Similarity:	22.93%	Mismatches:	179
Query Match:	8.98%	Indels:	113
DB:	9	Gaps:	17
US-10-066-551-4 (1-525) x US-09-738-626-3014 (1-1500)			
QY 70	ThrasnclyleuSeraspIleTyrlleuLeuAspProhIsaGlualPheAlaIaarg	89	
DB 346	ACCGGCGGAACAACAGGCTTAC-----TCGACACTACCGAATCCCTTAACGCGATG	399	
QY 90	AlaAlaIleuIleGluseralagIunIsSerleuAspIleuInTyrlIleTPArgAsn	109	
DB 400	ACCGCCCAATCGACGAAGCCGAAGAATACATCTAGTCGAGATCTACATCATGGCTGG	459	
QY 110	AspIleSerArgIaArgleuLeuPheAsnLeuValTyrlleuAlaIaIaGlualArg	129	
DB 460	GATTCTTACACCCACCATTTCTTCGACAGCATCGAACGACCCACACCGCGGCTCAA	519	
QY 130	ValaArgleuLeuAsp-----	135	
DB 520	GTCCGACTCTTTTCGACACAGTCGGCAGCTGGAATACCCGGCTACCACCGCTCAA	579	
QY 136	--AspaAsnThrArgglyleuAspAspIleuAlaIeuaSpSerHisproAsn	154	
DB 580	AAAGACACACCGCATGGCTTCGGCTGATCCATG-----	618	
QY 155	IleGlualArgleuPheAsnProPheValIleuAlaGlySTPArgAlaIeuglyTyrlleu	174	
DB 619	-----CTCCCTCCCAACCTGGCA-----	639	
QY 175	ThrasPheProaArgleuAsnArgArMeHisAsnIleSerPheThAlaAspAsnArg	194	
DB 640	CGCGGCTTCGCGCGACCCGACCTGCCAACACCGGAATGCTATCATGACGCGCAC	699	
QY 195	AlaThrIleuIleuglyIaArgAsn--IleGlyAspGluTyrl-----Phe	208	
DB 700	ACCGCATTCATCTCCCAAAATCTCATGGCCGAGTTACTTCACAAAAGAAAATC	759	
QY 209	LysValaIleGlyAspThrValPheAlaAspIleLeuAlaThgIlySerValVal	228	
DB 760	AAACTCGGCGCCGAA-----TGAAGAACCTCATGTGTAATCAGCGGCCCATCTGC	813	
QY 229	GlyIuValSerHisAspPheAspArgTyrlTPAlaSerHisSerAlaHisAsnAlaThr	248	
DB 814	TCCTCCATGAAATGATCTTCGCGGCGAGCTGTACGTGAATTCAC-----	861	
QY 249	ArgIleIleArgSerGlyAsnIleGlyGlyLeuGlnAlaIeuglyTyrlAsnAspGlu	268	
DB 862	-----GAAGCCCTCGACATCCGCGACAC	885	
QY 269	ThSerArgHisAlaIeuaArgTyrlArgGluThValGluGlnSerProleuTyrlGln	288	
DB 886	GCAGAGCCCGCTACATCGGCACACATCAAAAAGACTCCGCCACCAACCTCTGTCAG	945	
QY 289	LysIleGlnThrArgIleAspTrpGlnSerValGlnThrArgIleuIleSerAsp	308	
DB 946	CTCATCTCCCTCGCG-----	960	
QY 309	ProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyArgIleuInAsp	328	
DB 961	CCTGTTACACACAGAAACCAACCTGCG-----ATGTTACATCC	1002	
QY 329	AlaIeLysGlnProGlyLysSerValTyrlleuValSerProTyrlPheValProThrLys	348	
DB 1003	ATCGTTACACACGCAAGAAACATCATCTTGTGACGCCCTACTTCACTCCCGACGAA	1062	
QY 349	SerGlyThrAspAlaIeuaIaLysIleuValGlnAspGlyIleAspValThValIeuthr	368	
DB 1063	TCCTCTCTCGAAGCGTACCTACGCTGCTACCGGAGGATTAACCTGCAATCTATCTGTC	1122	

QY 369 AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLys 388  
DB 1123 TCT-----GAAACAAGCCCAATTCGCATCGACACAGCCCAATCCTCTACTACAG 1176  
QY 389 ProLeuLeuysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAla 408  
DB 1177 GCATCTCTTGAAGCCGGGTGAAATCTACCAATTC-----CCCAAAACCGAGCTC----- 1227  
QY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIle 428  
DB 1228 -----CTCCACACCAAGTAAATGATC 1248  
QY 429 ValAsp-----GlyLysArgIlePhe-----IleGlySerPheAsnLeu 441  
DB 1249 GCCGACCCCGACGACACACCGACGAGCCCTCGAGTCTCGATCTCTCCAACTC 1308  
QY 442 AspProArgSerAlaAlaGlyLeuAsnThrGluMetGlyValIleGluSerProLysIle 461  
DB 1309 GACATCGCGAGCTTGGCTCTCACTACGAAATCTCCTGATGATCGCCAAAGCGACCTC 1368  
QY 462 AlaGluGlnMetGlyArgThrLeuAlaAspThrSerProGluTyrAlaTyrArgValThr 481  
DB 1369 ATCCACGACATC---AACGCCCTCACCGACCGCTTACCGACAGTAAGTTTCAACTCAC 1425  
QY 482 LeuAspArgHisAsnArgLeuGlnThrPheHisAspProAlaThrArgLysThrTyr 499  
DB 1426 TTGGATTAAGTGAACGACGACGCGAGTTGG-----CGGCGCCGCTAC 1464

RESULT 10  
US-09-848-726-1  
Sequence 1, Application US/09848726  
Patent No. US2002010267A1  
GENERAL INFORMATION:  
APPLICANT: NAMPOOTHRI, Madhavan  
TITLE OF INVENTION: No. US2002010267A1el Nucleotide Sequences Coding for the cls ger  
FILE REFERENCE: 032301 WD 1191  
CURRENT APPLICATION NUMBER: US/09/848,726  
CURRENT FILING DATE: 2001-05-04  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1850  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (154)..(1653)  
OTHER INFORMATION:  
FEATURE:  
NAME/KEY: RBS  
LOCATION: (138)..(146)  
OTHER INFORMATION:  
US-09-848-726-1

Alignment Scores:  
Pred. No.: 1,49e-18 Length: 1850  
Score: 243.50 Matches: 105  
Percent Similarity: 36.24% Conservative: 61  
Best Local Similarity: 22.93% Mismatches: 179  
Query Match: 8.98% Indels: 113  
DB: 10 Gaps: 17

US-10-066-551-4 (1-525) x US-09-848-726-1 (1-1850)

QY 70 ThrAsnGlyLeuSerAspIleTyrLeuLeuAsnProHisGlnAlaPheAlaAlaArg 89  
DB 499 ACCGGCGGAACAACAGCTTCTAC-----TCCGACTACCGTGAATCCCTCAAAAGCATG 552  
QY 90 AlaAlaLeuIleGluSerAlaGlnHisSerLeuAspLeuGlnTyrTyrIleTyrArgAsn 109  
DB 553 ACCGGCCCAATCGACGAAAGCGGAAGATACATCTACGAGATCTCATCATGCGCTGG 612

QY 110 AspIleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgValArg 129  
DB 613 GATTCCTACACCAACCATCTCTTCGACAGCACTGACAGAGCCCAACCGGGGGTCAAA 672  
QY 130 ValArgLeuLeuLeuAsp----- 135  
DB 673 GTCCGACTCTCTTTCGACACAGCTGGAGCTGGAATACCCCGCTACACCGCTCAAA 732  
QY 136 ---AspAsnAsnThrArgGlyLeuAspPheLeuAspLeuLeuAlaLeuAspSerHisProAsn 154  
DB 733 AAGAACTCAACCCGATGGCTTGGCTGCTGACTCATG----- 771  
QY 155 IleGluValArgLeuPheAsnProPheValLeuArgLysThrArgAlaLeuGlyTyrLeu 174  
DB 772 -----CTCCCTCTCAACCTGGCGA----- 792  
QY 175 ThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArg 194  
DB 793 CGCCGCTTCGCGCCGACCCGACCTGCGACACCGCAAAATGCTCATCATGAGCGGCCAC 852  
QY 195 AlaThrIleLeuGlyArgAsn---IleGlyAspGluTyr-----Phe 208  
DB 853 ACCGCAATTCATGGCTCCCAAAATCTCATGCGCCCGAGTTACCTACAAAGAAAACATC 912  
QY 209 LysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySerValVal 228  
DB 913 AATCTCGCGCGCGAA-----TGGAAAGACCTCATGTGCTGACACTCAACGCGCCCATCTGC 966  
QY 229 GlyIleValSerHisAspPheAspArgTyrThrAlaSerHisSerAlaHisAsnAlaThr 248  
DB 967 TCTCTCATGGAATGATCTTCGCGCGAGCTGTGACTCGCAATTCAC----- 1014  
QY 249 ArgIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGlu 268  
DB 1015 -----GAACCCCTCGACATCCGCGACAC 1038  
QY 269 ThrSerArgHisAlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGln 288  
DB 1039 GCAGAGCCCAACGCTCATCGGACACACCTCAAAAGACCTCCGACCAACCTCTGTCAG 1098  
QY 289 LysIleGlnThrGlyArgIleAspThrGlnSerValGlnThrArgLeuIleSerAspAsp 308  
DB 1099 CTCATCCCTCCGCG----- 1113  
QY 309 ProAlaLysGlyLeuAspArgAspArgLysProIleAlaGlyArgLeuGlnAsp 328  
DB 1114 CTTGTTACACACACGAAACCAACCTGCG-----ATGTTCAACTCC 1155  
QY 329 AlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyrPheValProThrLys 348  
DB 1156 ATCGTTACACACGCGCAAGAACGACTCATCTTGTGACGCCCCCTACCTCATCCCGAGAA 1215  
QY 349 SerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThr 368  
DB 1216 TCCCTCTCGAAGCGCTGACCTGAGCTGACCGCGAGATGACGCTGAGACTATTGCTC 1275  
QY 369 AsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLys 388  
DB 1276 TCT-----GAAACAAGCCCAATTCGCATCGACACAGCCCAATCCTCTACTACAG 1329  
QY 389 ProLeuLeuysAlaGlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValProAla 408  
DB 1330 GCATCTCTTGAAGCCGGGTGAAATCTACCAATTC-----CCCAAAACCGAGCTC----- 1380  
QY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIle 428  
DB 1381 -----CTCCACACCAAGTAAATGATC 1401  
QY 429 ValAsp-----GlyLysArgIlePhe-----IleGlySerPheAsnLeu 441  
DB 1402 GCCGACCCCGACGACACACCGACGAGCCCTCGAGTCTCGGATCTCTCCAACTC 1461  
QY 442 AspProArgSerAlaAlaGlyLeuAsnThrGluMetGlyValIleGluSerProLysIle 461



Patent No. US20020127687A1  
GENERAL INFORMATION:  
APPLICANT: SHIGENOBU, SHUJI  
APPLICANT: MATANABE, HIDEKI  
APPLICANT: HATTORI, MASAHIRA  
APPLICANT: SAKAKI, YOSHIYUKI  
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS  
FILE REFERENCE: 081356/0159  
CURRENT APPLICATION NUMBER: US/09/790,988  
PRIORITY FILING DATE: 2001-02-23  
PRIORITY FILING DATE: 2000-04-07  
NUMBER OF SEQ ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 640681  
TYPE: DNA  
ORGANISM: Buchnera sp.  
US-09-790-988-1

Alignment Scores:  
Pred. No.: 2,87e-10 Length: 640681  
Score: 208.00 Matches: 105  
Percent Similarity: 37.88% Conservative: 81  
Best Local Similarity: 21.38% Mismatches: 163  
Query Match: 7.67% Indels: 142  
DB: Gaps: 21

US-10-066-551-4 (1-525) x US-09-790-988-1 (1-640681)

QY 49 HispHeasnthrSerLysProValArgLeu-----AspAsnIleLeuGlnIle 64  
DB 299339 CACGGAATTAAAGCAGAAATTAATTAACTAATACCAAGAAATTAATGCAATA 299300  
QY 65 ArgHsIthrProHsIthrAsnGlyLeuSerAspIleTyrLeuLeuAsnProHsIglu 84  
DB 299299 TTA-----ATACGTGATATTATTA----- 299279  
QY 85 AlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHsSerLeuAspLeuGlnTyr 104  
DB 299278 -----GCACGTAAAAATTAATGAAATGCTTTT 299252  
QY 105 TyrIleTparG-----AsnAspIleSerGlyArgLeuLeuPheAsnLeu 119  
DB 299251 TATATTGGAAACCGATGCGATGCGACATGATGCGCAATAGCTCTGATTTGAT----- 299198  
QY 120 ValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuAspAsnAsnThr 139  
DB 299197 -----TCTGCAAAACGTGATACATTGATTAATGCTGATTCAGCAGAGAT 299147  
QY 140 ArgGlyLeuAspAspLeuLeuLeuAlaLeuAspSerHsIProAsnIleGluValArgLeu 159  
DB 299146 -----ATTGAATTTTTCAGAGCCCTGCGTGAATTAATGAGA 299108  
QY 160 PheAsnProPheValLeuArgLysTyrParGAlaLeuGlyTyrLeuThrAspPhe---Pro 178  
DB 299107 AAATCTGGAATTCAGATGTAAGAGCTTTAAAGGAATTAATTAATGAGTTTAAAGA 299048  
QY 179 ArgLeuAsnArgMetHsIAsnLysSerPheThrAlaAspAspArgAlaThrIleLeu 198  
DB 299047 CGTGTGATGTTAGACACACAGAAATTAATTAATGATTAATTAATGATTCATCT 298988  
QY 199 GlyIleArgAsnIleGlyAspGlyTyrPhe-----LysValGlyGluAspThrValPhe 216  
DB 298987 GGTGACATGAACTGTGTGATCTTAATTAATTAATTAATTAATTAATTAATTAATG 298928  
QY 217 AlaAspLeuAspIleLeuAlaThrGlySerValAlaGlyGluValSerHsIAspPheAsp 236  
DB 298927 ATGATCTTAATGACAGAAATGAAAGGCGCTATGCTACACAAATGCGTAAATTAATCT 298868  
QY 237 ArgTyrTrpAlaSerHsIserAlaHsIAsnAlaThrArgIleIleArgSerGlyAsnIle 256  
DB 298867 TGTGATTCG-----GAAATT 298853

QY 257 GlyLysGlyLeuGlnAlaAlaLeuGlyTyrAsnAspGlyThrSerArgHsAlaLeuLeuArg 276  
DB 298852 GAACAGCTTTAAAAATCTTG----- 298832  
QY 277 TyrArgGlyThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAsp 296  
DB 298831 -----CCTCAATTACCAAAATTAATAAAAGTTAGAAATCAATCAATTAATAA 298787  
QY 297 TrpGlnSerValGlnThrArgIleIleSerAspAspProAlaLysGlyLeuAspArgAsp 316  
DB 298786 AACGCTGATTTCA-----GTAATTCATCCGACCGCGT----- 298751  
QY 317 ArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSer 336  
DB 298750 TTTCTTAATAAATATGATTCATCAACAGCGCTATTAAGACGATTTATAGCAAAACGCGAA 298691  
QY 337 ValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaLys 356  
DB 298690 TTAATCATAACTACCCCTTAATTAATGATCCAGATTTATTAAGAGCTAATTTGACT 298631  
QY 357 LeuValGlnAspGlyIleAspValThrValLeu-----ThrAsnSerLeuGln 372  
DB 298630 GCAGCTCAAAAGAGAGTAGAAGTATGATCATCTTAATACCTTAATATGATTCATTTTA 298571  
QY 373 AlaThrAspValAlaAlaAlaValHsIserGlyTyrValLysTyrArgLysProLeuLys 392  
DB 298570 GTAAATGCGCCAGTAGAGT-----TTTTTAGAGAGTATTAGAA 298529  
QY 393 AlaGlyIleLysLeuTyrGluLeuGlnProAsnHsAlaValProAlaThrLysAspLys 412  
DB 298528 CGAGCTGTAAATAATTTTCAATTTCA-----AAA 298499  
QY 413 GlyLeuThrGlySerSerValThrSerLeuHsAlaLysThrPheIleValAspGlyLys 432  
DB 298498 GGATTA-----TTACATGTAAGATATATAGATGACACAA 298460  
QY 433 ArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaAlaGluLeuAsnThrGluMet 452  
DB 298459 CTAAGTTAATTTGACAGCTTAATCTGATGATGAGAGTCTTTGTTGAATTTGGAATA 298400  
QY 453 GlyValAlaIleGluSerProLysIleAlaGluGlnMet-----GluArgThr 468  
DB 298399 ACTTGGTATTGATGATAGATGATTTGGACGCTAATTTGTTTGTATACAAATAATATAT 298340  
QY 469 LeuAlaAspThrSerProGluTyrAlaTyrArgValThrLeuAspArgHsIAsnArgLeu 488  
DB 298339 ATTTCTGATTCACAG-----TTAATAGATTAATAA----- 298310  
QY 489 GlnThrPheAspProAlaThrArgLysThrTyrProAsnGluProGluAlaLysLeuTrp 508  
DB 298309 GCTTGG-----TCTATGCGCGCATAT-----TGC 298286  
QY 509 LysArgIleAlaAlaLysIleLeuSerLeuLeu 519  
DB 298285 AAAGAAATTTTGAATAAATATTTTACTTTTA 298253

RESULT 13  
US-09-974-300-1093  
Sequence 1093, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berk, Randy M.  
APPLICANT: Clausen, Ib Groch  
TITLE OF INVENTION: Methods for Monitoring Multiple Gene  
FILE REFERENCE: 10085-500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
PRIORITY FILING DATE: 2001-10-05  
PRIORITY FILING DATE: 2000-10-06  
PRIORITY FILING DATE: 60/279,526  
PRIORITY FILING DATE: 2001-03-27

```

? NUMBER OF SEQ ID NOS: 8481
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 1093
? LENGTH: 933
? TYPE: DNA
? ORGANISM: Bacillus licheniformis
US-09-974-300-1093

```

Alignment Scores:	
Pred. No.:	1,41e-13
Score:	197.50
Percent Similarity:	34.99%
Best Local Similarity:	22.45%
Query Match:	7.28%
DB:	10
Length:	933
Matches:	86
Conservative:	48
Mismatches:	148
Indels:	107
Gaps:	11

US-10-066-551-4 (1-525) x US-09-974-300-1093 (1-933)

[illegible]

```

QY      396  LysLeuTYrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThr 415
          ::::|
Db      654  TATATTTCACCGTACTAC-----AGAGGCTTT--- 680
QY      416  GlySerSerValThrSerLeuHisAlaYsthrPheIleValAspGlyLysArgIlePhe 435
          |||||
Db      681  -----TATCATGCAAAAGCATTAATGCTGATGACCGGCGATGTGATG 722
QY      436  IleGlySerPheAsnLeuAspProArgSerIleAlaGlyLeuAsrThrGluMetGlyValVal 455
          |||||
Db      723  ATCGGAACATGCAATTTTGTACACAGAGACCGCTGTTCTCAAGTGAAGAACTGATGTCGTC 782
QY      456  IleGlySerProLys-----IleAlaGluGlnMetGluArg 467
          |||
Db      783  ATCCATGATTAAGACTGCACAAAGCAATCTTCGACGTCGTCAAGAAAGCATTTGACAC 842
QY      468  ThrLeuAlaAspThrSerProGluTyrAlaYrArgValThrLeuAspArgHisAsnArg 487
          |||
Db      843  GCCGAGCTTCGACAAAGAGGCGGTATGCGAAGCGCGCGGTATGATGACGCG-----CCC 896
QY      488  LeuGlnTTP 490
          ::::|
Db      897  GTCGAATG 905

RESULT 14
US-09-070-927A-415/c
Sequence 415, Application US/09070927A
Patent No. US20020120116A1
GENERAL INFORMATION:
APPLICANT: Charles A. Kunsch
          Patrick J. Dillon
          Steven Barash
          TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
          NUMBER OF SEQUENCES: 962
          CORRESPONDENCE ADDRESS:
          ADDRESSEE: Human Genome Sciences, Inc.
          STREET: 9410 Key West Avenue
          CITY: Rockville
          STATE: Maryland
          COUNTRY: USA
          ZIP: 20850
          COMPUTER READABLE FORM:
          MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
          COMPUTER: HP Vectra 486/33
          OPERATING SYSTEM: MSDOS version 6.2
          SOFTWARE: ASCII Text
          CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/09/070,927A
          FILING DATE: 04-May-2000
          CLASSIFICATION: <unknown>
          PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 60/046,655
          FILING DATE: 1997-05-16
          APPLICATION NUMBER: 60/044,031
          FILING DATE: 1997-05-06
          APPLICATION NUMBER: 60/066,009
          FILING DATE: 1997-11-14
          ATTORNEY/AGENT INFORMATION:
          NAME: Kenley K. Hoover
          REGISTRATION NUMBER: 40,302
          REFERENCE/DOCKET NUMBER: PB369
          TELECOMMUNICATION INFORMATION:
          TELEPHONE: (301) 309-8504
          TELEFAX: (301) 309-8512
          INFORMATION FOR SEQ ID NO: 415:
          SEQUENCE CHARACTERISTICS:
          LENGTH: 6285 base pairs
          TYPE: nucleic acid
          STRANDEDNESS: double
          TOPOLOGY: linear
          SEQUENCE DESCRIPTION: SEQ ID NO: 415:
US-09-070-927A-415

```

## Alignment Scores:

Pred. No.: 4.88e-08 Length: 6285  
 Score: 162.00 Matches: 52  
 Percent Similarity: 44.50% Conservative: 33  
 Best Local Similarity: 27.23% Mismatches: 69  
 Query Match: 5.97% Indels: 38  
 DB: 10 Gaps: 6

US-10-066-551-4 (1-525) x US-09-070-927A-415 (1-6285)

OY 287 TYRGLNLSLEGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgIleSer 306  
 Db 6122 TACCGAAGATC-----ATGAAGTGGAGCAT-TTCTCTTACAGCTTTTCA 6076  
 OY 307 ASPASPProAlaLysGlyLeuAspArgArgLysProIleAlaGlyArgLeu 326  
 Db 6075 GATGTCCT-----GATTCAGAAAGAAATTTGAAAGCGGTTT 6034  
 OY 327 GlnAspAlaLeuLysGlnProGlnLysSerValTyrLeuValSerProTyrPheValPro 346  
 Db 6033 GTTCGATGATTTCTCTCGCAAAAGAGTCTGTGATTCAAACACCGATTATTCCT 5974  
 OY 347 ThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrVal 366  
 Db 5973 GATGATGATGATTAATGCTTGTAGTGGCGGTTCTTCAGGTCGTCACGTCGCAAT 5914  
 OY 367 LeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyr 386  
 Db 5913 ATGATTCCTTGTATG-----CCTGACCAACCCCTTATTTACCGCGCACCAATATTAC 5860  
 OY 387 ArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406  
 Db 5839 GCAAAATTCCTTACACAAACCGCGCATTAATAATTAT----- 5824  
 OY 407 ProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThr 426  
 Db 5823 -----ATTATGATGATGAGGTT-----ATTCATCCAAACCA 5791  
 OY 427 PheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAla 446  
 Db 5790 ATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5731  
 OY 447 ArgLeuAsnThrGlnMeGlyValAlaIleGlnSerProLysIleAla----- 462  
 Db 5730 TCATTGAATTTTGGAGTGAAGTGAATTTTGAATCCGACATTCGCGATTTACCG 5671  
 OY 463 -----GluGlnMetGlnArgThrLeu 469  
 Db 5670 CAAGCTTTGAAGATATATGAAAAAAGTTC 5638

## RESULT 15

US-09-974-300-5529  
 : Sequence 5529, Application US/09974300  
 : Patent No. US20020146721A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Beika, Randy M.  
 : APPLICANT: Clausen, Ib Groth  
 : TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 : FILE REFERENCE: 10085-500-US  
 : CURRENT APPLICATION NUMBER: US/09/974,300  
 : PRIOR FILING DATE: 2001-10-05  
 : PRIOR APPLICATION NUMBER: 09/680,598  
 : PRIOR FILING DATE: 2000-10-06  
 : PRIOR APPLICATION NUMBER: 60/279,526  
 : PRIOR FILING DATE: 2001-03-27  
 : NUMBER OF SEQ ID NOS: 8481  
 : SOFTWARE: FASTSEQ for Windows Version 4.0  
 : SEQ ID NO 5529  
 : LENGTH: 738  
 : TYPE: DNA  
 : ORGANISM: Bacillus clausii

## Alignment Scores:

Pred. No.: 2.06e-09 Length: 738  
 Score: 161.00 Matches: 44  
 Percent Similarity: 47.73% Conservative: 19  
 Best Local Similarity: 33.33% Mismatches: 57  
 Query Match: 5.93% Indels: 12  
 DB: 10 Gaps: 2

US-10-066-551-4 (1-525) x US-09-974-300-5529 (1-738)

OY 78 LeuLeuAsnAspProHisGlnAlaPheAlaAlaAlaLeuIleGlnSerAlaGlu 97  
 Db 366 ATTTTCACGAGCGCAAAAGAAATTTGCGCATTTGTAAAGACATCAACAGCCGAG 425  
 OY 98 HisSerLeuAspLeuGlnTyrTyrIleThrPheAsnAspIleSerGlyArgLeuPhe 117  
 Db 426 ACGTTCATCCACATCCAGTATTAATTTCCGTTACGATGAGATCGAAGAACGATCGTG 485  
 OY 118 AsnLeuValTyrLeuAlaAlaGlnArgLysValArgValArgLeuLeuAspAsn 137  
 Db 486 GATCGCTGACAGAAAGCGCGCCAGCGTAAAGTCCGTTCTCTATGATGACCTC 545  
 OY 138 AsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluVal 157  
 Db 546 GGTTCGCGTCTTGTGCGCAAAAGATATTAAAGCGTTTCGTGAGACGCGGAGAAATTC 605  
 OY 158 ArgLeuPheAsnPro-----PheValLeuArgLysTyrPheArgAlaLeuGlyTyrLeuThr 175  
 Db 606 GGGATTTTTCCTCCATCGAGTTTACGATTTGTAACCTG----- 644  
 OY 176 AspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAla 195  
 Db 645 -----CGTTGAATTCCTCGCAACCAACCGTAACCTGTAACATGATGCGAAATGC 695  
 OY 196 ThrIleLeuGlyGlyArgAsnIleGlyAspGluTyr 207  
 Db 696 GGTATGTGCGGCGCTTACGTTGAGATGATGAT 731

## RESULT 16

US-09-938-842A-2104  
 : Sequence 2104, Application US/09938842A  
 : Patent No. US20020160378A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Harper, Jeff  
 : APPLICANT: Kieps, Joel  
 : APPLICANT: Wang, Xun  
 : APPLICANT: Zhu, Tong  
 : TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 : FILE REFERENCE: SAME, AND METHODS OF USE  
 : CURRENT APPLICATION NUMBER: US/09/938,842A  
 : PRIOR FILING DATE: 2001-08-24  
 : PRIOR APPLICATION NUMBER: US 60/227,866  
 : PRIOR FILING DATE: 2000-08-24  
 : PRIOR APPLICATION NUMBER: US 60/264,647  
 : PRIOR FILING DATE: 2001-01-16  
 : PRIOR APPLICATION NUMBER: US 60/300,111  
 : PRIOR FILING DATE: 2001-06-22  
 : NUMBER OF SEQ ID NOS: 5379  
 : SEQ ID NO 2104  
 : LENGTH: 3972  
 : TYPE: DNA  
 : ORGANISM: Arabidopsis thaliana  
 : US-09-938-842A-2104

## Alignment Scores:

Pred. No.: 0.00385 Length: 3972  
 Score: 118.00 Matches: 108  
 Percent Similarity: 34.44% Conservative: 68  
 Best Local Similarity: 21.14% Mismatches: 189  
 Query Match: 4.35% Indels: 146





```

QY 50 eAsnThSerlySproValArgLeuAspAsnIleuGlnIleArgHisThrProHisTh
Db 136 -----CCTGG-----CCTCACC 149
QY 70 rAsnGlyLeuSerSpleIleTyLeuLeuAsnAspProHisGlnAlaPheAlaAlaArgAl
Db 150 ACTTGCAAGAAAGAGATGATTAATTCATTCGATCATGACAG-----195
QY 90 aAlaLeuIleGluSerAlaGluHisSerLeuAspLeuGlnIleTyTyIleTArgAsnAs
Db 196 -----CAGATTA 203
QY 110 pIleSerGlyArgLeuLeuPheAsnLeuValTyLeuAlaAlaGluArgGlyValArgVa
Db 204 TAATCTCTTAAGATGATTTGCGATACCATT---CGCCAACTCAAGAGGATATTAT 260
QY 130 lArgLeuLeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeu--149
Db 261 GCGCATTTAC-----ACCATATCATCTGATGACATTTATCATCTCTAT 305
QY 150 -----AspSerHisProAsnIleGluValArgLeuPhe-----AsnProPh 163
Db 306 TCAGACTTCGACCATGTCCTGTAGAAAGTCAATACCATTCGCGAAGAAAGCTTACCTGT 365
QY 163 eValLeuArgIleTyArgAlaLeuGlyTyTyLeuThrAspPheProArgLeuAsnArgAr
Db 366 AGCATGTCAAAACTCGAGAGCTGCTCGCTGCTGACTAAGCAGAAAGCCCTC-----417
QY 183 gMetHisAsnIleSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArgAsnI
Db 418 -CAACATTAATAAAACTGATTTGGCTGATTTCCAAACAGTATACAGGATCAGCCAAC--474
QY 203 eGlyAspGluTyTyPheLeuValGlyGluAspThrValPheAlaAspLeuAspIleLeuAl
Db 475 -----TACACGACTTGCTGCTC-----492
QY 223 aThrGlySerValValGlyGluValSerHisAspPheAspArgTyTyTrpAlaSerHisSe
Db 493 -----AATCAGAT-----501
QY 243 rAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGlyGluGlnAlaLe
Db 502 -----GCCACGCTACGCGCATGATA-----522
QY 263 uGlyTyAsnAspGluThrSerArgHisAlaLeuLeuArgTyArgIleuThrValGluG
Db 523 -----GAAAG 527
QY 283 nSerProLeuTyGlnIleGlnIleThrGlyArgIleAspTrpGlnSerValGlnThrAr
Db 528 TTTCGAATTTACATGACGACGCTTTTCTGAAGACCCCAACCTGTCATGCGGACCTCA 587
QY 303 gLeuIleSerAspAspProAlaIleGlyLeuAspArgAspArgGlyGlySProPheIleAl
Db 588 GCTGCTCAATTTACTTCTATCCAG-----CGTTGATTCCTATATGACAGC 632
QY 323 aGlyArg--LeuGlnAspAlaLeuGlyGlnProGluIleSerValTyTyLeu-----339
Db 633 ATCAAAATGATTTTGAATGCAATTAACCAAGCAAGCAGCATTTTGTGCTTGATGTA 692
QY 340 -----ValSerProTyTyPheValProThrIleSerGlyThrAspAlaLeuAlaIleSle
Db 693 TATCTTCTTAAGCCCAATTTCTCTTA-----GCTCTTGCCCAAGC 734
QY 357 uValGlnAspGlyLeuAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAl
Db 735 TATCGAAGAGAGATGTGAGTAAGTAATCATGACACATTCACAAACAGATACATG 794
QY 377 aAlaValHisSerGlyTyTyValIleTyTyArgGlySProLeuLeuIleGlyIleTySle
Db 795 C-----AACTACTGAGCAATTTGGTATACCACT 824

```

```

QY 397 u-----TyGluLeuGlnProAsnHisAlaValProAlaThrIleAspIleGlyLeuTh
Db 825 TCCTATTACGAA-----AGAAAAAGGAGGCGTT--855
QY 415 rGlySerSerValThrSerLeuHisAlaIleTyTyPheIleValAspGlyIleAspIlePh
Db 856 -----CTCCATTAAGATTTGTTGATCGACATTAATACTTAAT 896
QY 435 eIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyVala
Db 897 CTTCGCTCTGCTACAGCGGAGCGGCTGCTGATCATTAATAAAGCTTGA---GACCTATT 953
QY 455 lIleGluSerProIleAlaGluGlnMetGluArg-----ThrLeuAl 470
Db 954 CATCTTCGCGCAATTAACAGACACAGCTTACAGCCCTTATGACGCTGCTCTCTCT 1013
QY 470 aAspTrSerProGluTyTyValTyTyArgValThrLeuAspArgHisAsnIleGluGlnI
Db 1014 AGAACAATAAGCTCTCTAT-----CTGTCCACAGAGAGCGTCTTACGCG 1058
QY 490 pHisAspProAlaThrArgIleGlyThr 498
Db 1059 TCCTACTCTCTCAAGTAGACCTACT 1083

RESULT 18
US-09-938-842A-2045
; Sequence 2045, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938, 842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227, 866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264, 647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300, 111
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2045
; LENGTH: 3612
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2045

Alignment Scores:
Pred. No.: 0.017 Length: 3612
Score: 112.00 Matches: 93
Percent Similarity: 37.50% Conservative: 57
Best Local Similarity: 23.25% Mismatches: 148
Query Match: 4.13% Indels: 103
Gaps: 20

US-10-066-551-4 (1-525) x US-09-938-842A-2045 (1-3612)
QY 109 AsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyTyLeuAlaAlaGluArgGlyVal 128
Db 1328 AATGTTCTATCGCGGAGCAAGCTTATGACGCTGATTAACGTAA-GTTGAAGATCTATG 1386
QY 129 ArgValArgLeuLeuLeuAspAsnAsnThrArgIleTyTy-----AspAspLeuLeu 146
Db 1387 GCTATGAGA-----AGCAATGCTTCGAGTTTGCTGCTGATGATTTGGAT 1431
QY 147 LeuAlaLeuAsp--SerHisProAsnIleGlu-----156
Db 1432 GTTGTAGTGAACGAGCTGCTGCTCAAGATCAAAAGGTTTAATTTCTCGATGAGAGGTT 1491

```

```

Oy 157 -----ValArgLeuPheAsnProPheValLeuArgLys---TyrPargala 170
Db 1492 ATGAAGGAATGGGTATAGACGAGATGCGAGATTTTCACAAAGTTTTCAGGTTG 1551
Oy 171 LeuGlyTyrLeu---ThrAspPheProArgLeuAsnArgArgMethHisAsnLysSerPhe 189
Db 1552 TTAGCTGTTTTCACACGCGCATCTGAAACCGATGAGATGAGGAGGAGTCTCTTAT 1611
Oy 190 ThrAla-----AspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlu 206
Db 1612 GAGGCTGACTCCTCGATGAGAACTGCTTTGTTGCGGCTCGAGAAATTCGGGTTTGAG 1671
Oy 207 TyrPheLysValGlyGluAspThrVal---PheAlaAspLeuAspIleLeuValThrGly 225
Db 1672 TTCCTTACCGGACGCAAAACGCGATATCTTCCGTGAATTGAT---CTTCTCCAGGG 1728
Oy 226 SerValValGlyGluValSerHisAspPheAspArgTyrTyrPalaSerHisSerAlaHis 245
Db 1729 AAAACAGTTGAAAGATATATCGTTTAAACGTTTGAAGTTTAAACGCGCGAGAAAG 1788
Oy 246 AsnAlaThrArgIleIleArgSerGlyAsn-----IleGlyLysGlyLeu 260
Db 1789 AGAATGTCGTTATTTGTCGCGACAGATGGAGACTTCTGTTACTGTCTCAAGGAGCT 1848
Oy 261 GluAlaLeu----- 263
Db 1849 GATAATGTGATGTTGAAAGACTAGCCAAAAACGGGAGAAAGTTGAGAGAGAACTCGA 1908
Oy 264 -----GlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgTyrArgGlu 279
Db 1909 GAACATGTAAATGAAATATGATGATGATGAGGCGTAAAGACATGATCTTGCATACCGTGAG 1968
Oy 280 ThrValGluInSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspPrgInser 299
Db 1969 GTTGATGAGAT-----GAGTAT 1986
Oy 300 ValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgArgLys 319
Db 1987 ATAGAAATTCAGCAAGAACTTCATGAAGTAAAGTACGCTGCGGATTCGTGAG--- 2043
Oy 320 ProProlIleArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeu 339
Db 2044 -----AGCTTAATAGATGAATATACAGAACAGATGACGATTTGATTTCTC 2091
Oy 340 ValSerProTyrPheValProThrLys-----SerGlyTyr---AspAlaLeuAlaLys 356
Db 2092 CTGTGCTACTGCGAGTTCGAGACAACTTCAAATGGGGTTCGCGATTTGTTCACAGAG 2151
Oy 357 LeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspVal 376
Db 2152 CTAGCTCAAGCAGGAGATTAAAGATTGGGTTTACCCGAGACAAATGAGACT----- 2205
Oy 377 AlaAlaValHisSerGlyTyr----- 383
Db 2206 ---GCAATCAATATGATGATTTCTGTAGTTACTAAGACAAGATGAGACGATCATC 2262
Oy 384 -----ValLysTyrArgLysProLeuLysAlaGlyIleLys---LeuTyr 398
Db 2263 ATAAATCTTGAGACACACATATCAAGATTCGAGAAAGCTGGAGAGAAAGATGCAATT 2322
Oy 399 GluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly-----LeuThr 415
Db 2323 GAACATGCATCAAGAAAGAGCGTGTGTAACCAATGAGAGAAAGGAAACCTCTATCTACA 2382
Oy 416 GlySerSerValThrSerLeuHis---AlaLysThrPheIleValAlaLysGlyLysArgIle 434
Db 2383 GCATCAAGCAGCGCAAGCTCATCATGAGCCCTTGGCTTGCATTCATGACGGAAGTCGCTA 2442

```

RESULT 19  
 US-09-938-842A-1375  
 ; Sequence 1375, Application US/09938842A  
 ; Patent No. US20020160378A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1375
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1375

Alignment Scores:
Pred. No.: 0.0137
Score: 109.50
Percent Similarity: 34.41%
Best Local Similarity: 19.77%
Query Match: 4.04%
DB: 9
Gaps: 28

US-10-066-551-4 (1-525) x US-09-938-842A-1375 (1-2061)
Oy 11 ProserGluThrIleSerLeuMetLysThrArgSerIleIleSerLeuLysLeu 30
Db 58 CCATCGAATCGATCTCTTA-----ACCACCTGCTGACTACTTTTACAACTC 108
Oy 31 LeucySerCySerSerTyrLeuProLeuGluGluArgThrGluSerArgHisPhe 50
Db 109 GCCGCGAGATCTTCAAGCTTCAACCA-----AACACTTC 144
Oy 51 AsnThrSerLysProValArgLeuAspAsnIle----- 61
Db 145 TCTACCAACAAAGAAAGTTTAAAGAACTCCGACATGTTCAAACTCTGTGATCTTC 204
Oy 62 -----LeuGlnIleArgHisThrProHisThrAsnGlyLeuSerAspIle 76
Db 205 TTGTAAGACCTCCGATTAATATCCGAGTCGGGTCAATACCGCGGGTCTTGGTAATC 264
Oy 77 TyrLeuLeuAsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSer 95
Db 265 TTGAGCCTTTCAGAGCTTCACGTCACTTCGAGAAACTCAATCTCTACTACAGACTGT 324
Oy 96 -----AlaGluHisSer 99
Db 325 ACAAGATGAGCTAAGCTATATATGTTAATGAACCTCCGTCAGAGTCCGCTCATTTTC 384
Oy 100 LeuAspLeuGlnTyrTyrIleTyrPargAsn-----Asp 110
Db 385 CGGAGTTTGACCCGATCCATATTCACCTACCTCGACACCTTTCCAGCTAGCTGTGAG 444
Oy 111 IleSerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal 130
Db 445 TTACCTGGAGAA---GTTAATGAGCTAATCTTCTGTATATGCGTCAGACTCGAAATTC 501
Oy 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeuAspAspLeu----- 145
Db 502 GAAGCAAGACCCGACCGACGACACCAACAGCAAGCTATGAGCTGTATTATGTTCTTAAT 561
Oy 146 LeuLeuAlaLeuAspSerHisProAsnIleGlu-----ValArgLeuPheAsnProPhe 163
Db 562 CTGTTTGACAAACAGATTATTCGAAATTCCTGATGAATTCGAGATCTTATCATCAT 621
Oy 164 ValLeuArgLysTyrPargAlaLeuGlyTyrLeuThrAspPheProArgLeuAsnArgArg 183

```

```

Db 622 GGTGTAAGAAATGAGAGATTGTGTCAAGAGATCGATTTTTCAGAGAGATCTCC 681
      ::::::::::::::::::::
      MethisAsnLysSer-----
      184 ::::::::::::::::::::
      682 GTGGGGAAGAAAGATACATCGAATCGAGCTTCTCAGCAACTTAATGCGGCTTATGT 741
      ::::::::::::::::::::
      192 AspaSnArGAlaThrIleLeuGlyGlyArgAsnIleGlyAspGlyTyrPheLysValGly 211
      ::::::::::::::::::::
      742 TACTGCAGATGTGTATCTTCAGAGCATGTGATGTATGAAGAGAACAGAGAA 801
      ::::::::::::::::::::
      212 GluAspThrValPhe-----
      802 GAAGATGATTTTGATGAGTTCGAGTTTAAACGTGATGATCTTCGTCGATTTCT 861
      ::::::::::::::::::::
      219 LeuAspIleLeuAlaThrGlySerValValGlyValLysSerHisAspPheArgTyr 238
      ::::::::::::::::::::
      862 CTAGAGATCAAG---AGTATCATCCAGTGTGTTAGATACAGACACACTTAATGATCG--- 915
      ::::::::::::::::::::
      239 TrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIle----- 256
      ::::::::::::::::::::
      916 ---ACCTTATCAAAAGTGTTCCGCTCCGGAACATTACGTGC 957
      ::::::::::::::::::::
      257 ---GlyLysGlyLeuGlnAlaLeuGlyTyrAsnAspGlyThrSerArgHisAla 273
      ::::::::::::::::::::
      958 CCTAAACCGGTAAACTTGTGTTAGTCTGTGTTGCGTAACCTCTCT----- 1008
      ::::::::::::::::::::
      274 LeuLeuArgTyrArgGlyThrValGlyLysSerProLeuTyrGlnLysIleGlnThrGly 293
      ::::::::::::::::::::
      1009 ---GTCAGCAAGTATTCAAAGC-----TACTCCAAAC---CAAAAGCGT 1047
      ::::::::::::::::::::
      294 ArgIleAspTrpGlnSerValGlnThrArgLeuIleSerAspPheProAlaLysGlyLeu 313
      ::::::::::::::::::::
      1048 GTACGTTTATGCTCAAAAGCGAGAGAAAGAGTGTG---GATGTCGCGAGAGCTTTA 1098
      ::::::::::::::::::::
      314 AsparGAspArgArgLysProProIleAlaGlyArgLeuGln----- 327
      ::::::::::::::::::::
      1099 GCGGCGGAAGA-----GCTGGAACACTAACGCGGAGCTTCTCGCCGGA 1143
      ::::::::::::::::::::
      328 AspaLeuLysGlnProGlyLysSerValTyrLeuValSerProTyrPheValProThr 347
      ::::::::::::::::::::
      1144 GAATTGATCAAGGAGAGAGAGAA----- 1167
      ::::::::::::::::::::
      348 LysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeu 367
      ::::::::::::::::::::
      1168 ---GAGTGTGTTAAAGCATTAGTGGAGATTCGAATTCCT 1203
      ::::::::::::::::::::
      368 ThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArg 387
      ::::::::::::::::::::
      1204 ACCAAGAC-----AGACTTTTACAGA 1227
      ::::::::::::::::::::
      388 LysProLeuLeuLysAlaGlyIleLysLeuTyrGlyLeuGlnProAsnHisAlaValPro 407
      ::::::::::::::::::::
      1228 TCGGTTTAGTGAAGCGTGAAGT-----GTTGAATCGTTGATGAGATTCCTT 1275
      ::::::::::::::::::::
      408 AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPhe 427
      ::::::::::::::::::::
      1276 CGTTCCGATGATCCGAGATCCAGAGAACCCCATGCTGGCATTAATTCAGAA 1335
      ::::::::::::::::::::
      428 IleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArg 447
      ::::::::::::::::::::
      1336 GATATCCCGGAAAAGTCAATGCTCGAGAA-----GACGCGTGGATTAAAGG 1386
      ::::::::::::::::::::
      448 LeuAsnThrGluMetGlyValValIleGluSerProLysIleAlaGluMetGluArg 467
      ::::::::::::::::::::
      1387 CTA-----ATCGTTGAGGTTTAAAGAT 1410
      ::::::::::::::::::::
      468 ThrLeuAlaAspThrSerProGlyTyrAla-----TyrArgValThrLeu 482
      ::::::::::::::::::::
      1411 GGTGCGAAGAGAGACAGACATACGACACACCTGCTGTGTTTATCATTTCTCTC 1470
      ::::::::::::::::::::
      483 AsparGHisAsnArgLeu----- 488
      ::::::::::::::::::::

```

```

Db 1471 GGAGATTACACAGATTGATCGGAAATCTCTGATGCGATTCCGGATTGGTGAATC 1530
      ::::::::::::::::::::
      489 -----GlnTrpHisAspProAlaThrArgLysThr----- 498
      ::::::::::::::::::::
      1531 GTTAAAGTTCGATTTACGTTGATTCACGGAACGCAAGCGTTAATCGCATTCGAGCT 1590
      ::::::::::::::::::::
      499 ---TyrProAsnGluProGlyAlaLysLeuThrPysArgIleAlaAlaLysIleLeuSer 517
      ::::::::::::::::::::
      1591 TTGTTAATGAAACCAACCTGAT---AATCATCGGGGATCTCGCGCGGAATTTGCTT 1647
      ::::::::::::::::::::
      518 LeuLeu 519
      ::::::::::::::::::::
      1648 GTTCTT 1653
      ::::::::::::::::::::
      RESULT 20
      US-09-974-300-5542
      : Sequence 5542, Application US/09974300
      : Patent No. US20020146721A1
      : GENERAL INFORMATION:
      : APPLICANT: Berka, Randy M.
      : TITLE OF INVENTION: Methods For Monitoring Multiple Gene
      : FILE REFERENCE: 10085-500-US
      : CURRENT APPLICATION NUMBER: US/09/974,300
      : PRIOR FILING DATE: 2001-10-05
      : PRIOR APPLICATION NUMBER: 09/680,598
      : PRIOR FILING DATE: 2000-10-06
      : PRIOR APPLICATION NUMBER: 60/279,526
      : PRIOR FILING DATE: 2001-03-27
      : NUMBER OF SEQ ID NOS: 8481
      : SOFTWARE: FASTSEQ for Windows Version 4.0
      : SEQ ID NO 5542
      : LENGTH: 465
      : TYPE: DNA
      : ORGANISM: Bacillus clausii
      : FEATURE:
      : NAME/KEY: misc_feature
      : LOCATION: (1)..(465)
      : OTHER INFORMATION: n = A,T,C or G
      : US-09-974-300-5542
      :
      Alignment Scores:
      Pred. No.: 0.00144 Length: 465
      Score: 109.00 Matches: 34
      Percent Similarity: 41.77% Conservative: 32
      Best Local Similarity: 21.52% Mismatches: 58
      Query Match: 4.02% Indels: 34
      DB: 10 Gaps: 5
      :
      US-10-066-551-4 (1-525) x US-09-974-300-5542 (1-465)
      :
      287 TyrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThrArgLeuIleSer 306
      ::::::::::::::::::::
      3 TYCCACAGATTACAGACGAGCAACCAACCATTCGAGTTTATACACA----- 50
      ::::::::::::::::::::
      307 AspaSpProAlaLysGlyLeuAspArgAspArgLysProProIleAlaGlyArgLeu 326
      ::::::::::::::::::::
      51 -----CAAGGAGACGAGCTTGAACAA-----ACACTG 77
      ::::::::::::::::::::
      327 GlnAspAlaLeuLysGlnProGlyLysSerValTyrLeuValLysSerProTyrPheValPro 346
      ::::::::::::::::::::
      78 CTTATGCTTATTCCTACCGCTAAACAGCTTAATCATTCGCGACCCCTTATTTATCCCG 137
      ::::::::::::::::::::
      347 ThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrVal 366
      ::::::::::::::::::::
      138 ACACGGCGCGTAATGATGTCATATGAAGTGGCGAAGAGCGTCTTCCTTCTTA 197
      ::::::::::::::::::::
      367 LeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyr 386
      ::::::::::::::::::::
      198 ATCCCTCCT-----GAAAAAGCAGATCATCCCTTCTTCAACAGCGGCGTACCTTAT 251
      ::::::::::::::::::::
      387 ArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGlyLeuGlnProAsnHisAlaVal 406
      ::::::::::::::::::::

```

```

Db 252 GTGATTGAAGCCCTGACAGCAGCGTCGTTTACATTATTAT----- 296
QY 407 ProAlaThrLysAspLysArgLysSerSerValThrSerLeuHisAlaLysThr 426
Db 297 -----CGAGGATTT-----TACCATGTTAAAGCC 320
QY 427 PheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArg 444
Db 321 GTAGTAGCTGATCATCTGCATGTTTAGTGGGACCCCAATATCATAAAGCC 374

RESULT 21
US-10-033-297-21
: Sequence 21, Application US/10033297
: Publication No. US20020187486A1
GENERAL INFORMATION:
: APPLICANT: Hall, Jeff G.
: Lyamichev, Victor I.
: Mast, Andrea L.
: Brow, Mary Ann D.
TITLE OF INVENTION: Detection of Nucleic Acids By Multiple
: Sequential Invasive Cleavages
NUMBER OF SEQUENCES: 163
CORRESPONDENCE ADDRESS:
: ADDRESSEE: Medien & Carroll, LLP
: STREET: 220 Montgomery Street, Suite 2200
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94104
COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/10/033,297
: FILING DATE: 12-NOV-2001
: CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/09/350,597
: FILING DATE: 09-JUL-1999
: APPLICATION NUMBER: US/08/823,516
: FILING DATE: 24-MAR-1997
: APPLICATION NUMBER: PCT/US97/01072
: FILING DATE: 21-JAN-1997
: APPLICATION NUMBER: US 08/759,038
: FILING DATE: 02-DEC-1996
: APPLICATION NUMBER: US 08/758,314
: FILING DATE: 02-DEC-1996
: APPLICATION NUMBER: US 08/756,386
: FILING DATE: 29-NOV-1996
: APPLICATION NUMBER: US 08/682,853
: FILING DATE: 12-JUL-1996
: APPLICATION NUMBER: US 08/599,491
: FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
: NAME: Ingolia, Diane E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2502 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-033-297-21

```

```

Alignment Scores:
Pred. No.: 0.0486 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.708 Conservative: 65
Best Local Similarity: 21.798 Mismatches: 174
Query Match: 3.91% Indels: 188
DB: 9 Gaps: 30

US-10-066-551-4 (1-525) x US-10-033-297-21 (1-2502)
QY 118 AsnLeuValTyrLeuAlaIleGluArgGlyValArgVal----- 130
Db 961 GATCTTCTGCGCCCTGCGCCCGCCGAGGGGGGGGCGGTCACCGGCGCCCGACGCTTAT 1020
QY 131 ArgLeuLeuLeuAspAspAsnThrArgGlyLeu-----AspAspLeuLeu 147
Db 1021 AAGGCCCTCAGGACCTGAGAGAGCGGGGGCTTCTGCGCAAAACCTGACGCTTCG 1080
QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGAGCGCTTGGCCCTCCCGCGGAGAGACCCCATGCTC----- 1128
QY 164 ValLeuArgLysTyrPargAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTCGCCTACCTCTGTGACCTTCCACCAACACACCGCCGAG 1167
QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGGCCCGCGCTACGCGGGGAGTGAGCAGAGAGCGGGGAGCGGGCGCCCTT 1227
QY 198 -----LeuGlyArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGGCTCTTGCACAACTGTGGGGAGGCTTGAGAGGGAGAGAGGCTCTTGG 1287
QY 207 ---TyrPheLysValGlyLysLys-----AspThrValPheAlaAspLeuAspIleAla 223
Db 1288 CTTTACGGGAGGTGAGAGGCGCCCTTCCGCTGCTGCTGCGCCACATGAG-----GCC 1341
QY 224 ThrGlyLys-----SerValAlaGlyLys 230
Db 1342 ACCGGGGTCCGCTGAGCGTGCCTATCTCAGGGCCTTGTCCCTGAGAGGTGGCGGGAG 1401
QY 231 ValSer-----HisAspPheAspArgTyrTyr 239
Db 1402 ATCGCCGCGCTGAGCGCGAGGCTTCCGCTGCGCGCGGCGCACCCCTTCAAC----- 1452
QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259
Db 1453 ---CTCACTCCCGGAGCAGCTGGAAGAGGCTCTTGAC-----GAGCTAAGG 1500
QY 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
Db 1501 CTTCCCGCATCGCAAGACGAGAGACCGGAGCGCTCCACCGCGCGCGCTCTG 1560
QY 274 -----LeuLeuArgTyrArgGluThrVal 281
Db 1561 GAGCGCCCTCCGCGAGCGCCACCCCATCTGAGAGAAATCTCTGACGAGCGGAGCTCACC 1620
QY 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294
Db 1621 AAGCTGAAAGACACCTACATTGACCCCTTCCGCGACCTCATCACCACCGAGCGCGCG 1680
QY 295 IleAspTyrGlnSerValGlnThr-----ArgLeuIleSerAspPro 309
Db 1681 CTCACACCGCGCTTCAACACGAGCGCCACGCGGAGGAGGCTAAGTACCTCCGATGCC 1740
QY 310 AlaLysGlyLeuAspArgAspArgLysProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCCAGAAATCCCGCGCGCGACCGCGCTTGGAGAGATCCCGCGGGCC 1794
QY 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCATCGCCGAGAGGGGTGCTATTGTGGCTGACTATAGCCAGATAGAGTCAAG 1854

```

```

Oy 344 PheValProThrIlysserGlyThrAspAlaLeuAlaIleuValGlnAspGlyIleAsp 363
      ::::: ||||| ::::: ||| ::::: |||::: |||::: |||
Db 1855 GTGCTGGCCCAACCTCTCGCGGCAGACAGAACTGATCCGGGTCTTCAGAGAGGGCGGGAC 1914
Oy 364 ValThrValIleuThrAsnSer-----LeuGlnAlaThrAspValAla 377
      ::::: ||| ||| :::::
Db 1915 ATCCACACGAGAGACCCGACCTGATGTCGGCTCCCCGGAGAGCCGTGGAC----- 1968
Oy 378 AlaValHisserGlyThrValIlystyrGlyProLeuLeuIlyAla----- 393
      ::::: |||||:::
Db 1969 -----CCCTATGCGCGGGCGCCAGAAC 1995
Oy 394 -----GlyIleIlyLeuThrGlyLeuGlnProAsnHisAlaVal 406
      ::::: |||||::: |||::: |||:::
Db 1996 ATCAACTTCGGGGGCTCTACAGGCATGTGGCGCCACGGCTCTCCAGAGAGCTTACCATTC 2055
Oy 407 Pro----- 407
      |||
Db 2056 CCTTACGAGAGAGGCCAGGCTTCATTGAGCGCTACTTTCAGAGCTTCCCAAGCTGGCG 2115
Oy 408 -----AlaThrIlyAspIlyGlyLeuThrGlySerSerValThrSerIleuHis 423
      ||| ::::: |||||
Db 2116 GCCTGCAATTGAGAAGACCTCGGAGGAGGAGAGAGCGGGGCTAGCTGAGACCTCTTC 2175
Oy 424 AlaIlyThrPheIleValAspGlyIlystyrGlyIlePheIleGlySerPheAsnIleuAspPro 443
      |||||::: |||:::
Db 2176 -----GGCCCGCGCGCTACGTGCTCCAGACTGAGAGCCCGGGTG 2214
Oy 444 ArgSerAlaIleuAsnThrIleuMetGlyValAlaIleGlySerPro-----Lys 460
      ::::: ||| ||| :::::
Db 2215 AAGAGCTGGGGGAGGCGCGCCAGGCC--ATGGCTTCAGACATGCTCCCGGGGCAC 2271
Oy 461 IleAlaGlnMetGluArgThrIleuAlaAspThrSerPro-----GlyTyrAla 477
      ||||: ||||: ||||:
Db 2272 GCCCGCAGCTCATGAGCTGCGGTATGTCGTAACCTTCCACAGCTGAGAGAAATGGGG 2331
Oy 478 TyrArgValThrIleuAspArgHisAsnArgLeuGlnThrPheIlyAspProAlaThrArgLys 497
      ||||: ||| ||||: ||| |||
Db 2332 GCCAGGATGCTCTTCAGCTCCAGTCACGAGAGCTGCTCAGAGGCCCAAAAGAGGGCG 2391
Oy 498 ThrIlyProAsnGluProGluAlaIlyIleuThrPlyArgIleAlaAlaIlyIleuSer 517
      ||| |||
Db 2392 -----GAGCGCGCTGGCGCCGGCTGGCCAAAGAGGTATGAGGGGTGTATCC 2438
Oy 518 Leu---LeuProIleGlu 522
      ||| ::::: ||||
Db 2440 CTGGCGTGGCCCTGGAG 2457

RESULT 22
US-10-081-806-21
; Sequence 21, Application US/10081806
; Publication No. US20020197623A1
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
Hall, Jeff G.
Lyamichnev, Victor I.
TITLE OF INVENTION: Invasive Cleavage Of Nucleic Acids
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/081,806

```

```

?      FILING DATE: 22-Feb-2002
?      CLASSIFICATION: <Unknown>
?      PRIOR APPLICATION DATA:
?      APPLICATION NUMBER: US/08/756,386
?      FILING DATE: <Unknown>
?      APPLICATION NUMBER: US 08/682,853
?      FILING DATE: 12-JUL-1996
?      APPLICATION NUMBER: US 08/599,491
?      FILING DATE: 24-JAN-1996
?      ATTORNEY/AGENT INFORMATION:
?      NAME: Ingolia, Diane F.
?      REGISTRATION NUMBER: 40,027
?      REFERENCE/DOCKET NUMBER: FORS-02564
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (415) 705-8410
?      TELEFAX: (415) 397-8338
?      INFORMATION FOR SEQ ID NO: 21:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 2502 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: double
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      SEQUENCE DESCRIPTION: SEQ ID NO: 21:
?      US-10-081-806-21
?
?
?      Alignment Scores:
?      Prid. No.: 0.0486 Length: 2502
?      Score: 106.00 Matches: 119
?      Percent Similarity: 33.70% Conservative: 65
?      Best Local Similarity: 21.79% Mismatches: 174
?      Query Match: 3.91% Gaps: 188
?      DB: 9
?
?      US-10-066-551-4 (1-525) x US-10-081-806-21 (1-2502)
?
?      Oy 118 AsnLeuValTyrLeuAlaIaIaGluArgGlyValArgVal----- 130
?      ::::: ||| ||| ||| ||| ||| |||
?      Db 961 GATCTTCTGGCCCTGCGCCGCGCCGACGAGGGGGCGGCTCACCGGGCCCCCGAGCCTTAT 1020
?      ArgLeuLeuAspAspAsnAsnThrArgGlyLeu-----AspAspLeuLeu 147
?      :::: ||| ||| ||| ||| ||| ||| |||
?      Db 1021 AAAACCCCTCAGGACCTGAGAGGAGGGCGGCGCTTCGCCAAAGACTGACCTGACCTCG 1080
?      AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
?      ||||| ||| ||| ||| ||| |||
?      Db 1081 GCCCTGAGGAGGAGCCTTGCCCTCCCGCCGCGGACGACCCCATGCTC----- 1120
?      ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
?      164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
?      1129 -----CTCGCCCTACCTCTCTGACCTTCCAAACACACCCCGAG 1167
?      Oy 180 LeuAsnArgArgmethHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
?      1168 GGGGTGGCCCGGCGCTACGGCGGGGAGGTGACGAGAGGCGGGGAGCGGCCCGCTT 1227
?      198 -----LeuGlyGlyArgAsnIleGlyAspGlu----- 206
?      1228 TCCGAGAGGCTCTTCGCCCAACCTGTGGGGGAGGCTTGAAGGGGAGAGAGCTCTTGG 1287
?      TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
?      207 ---TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
?      1288 CTTTACCGGAGGTGAGAGGAGCCCTTTCGCTGCTGTCGACCCACATGGAG-----GCC 1341
?      224 ThrGly-----SerValValGlyGlu 230
?      1342 ACGGGGGTGGCCTTGACAGTGGCTTATCTCAGGGCCTTCCCTGGAGAGTGGCGGGGAG 1401
?      231 ValSer-----HisAspPheAspArgTyrTrp 239
?      1402 ATCCCGCGCCTCGAGGGCGGAGGCTTCCGCGTGGCGGCGGACCCACTTCAAC----- 1452
?      240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGly 259

```

```

Db 1453 ---CTCACTCCCGGAGCAGCTGGAAGGGCTCTTGTGAC-----GAGCTAGGG 1500
               |||
Oy 260 LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla----- 273
               |||
Db 1501 CTTCCCGCATCGGCAAGAGGAGAACCGGCAAGCGCTCCACAGCGCCCGCTCTG 1560
               |||
Oy 274 -----LeuLeuArgTyrArgLysThrVal 281
               |||
Db 1561 GAGGCGCTCCGGAGGAGCCCGCCATCGTGGAAGATCTTCAGTACCGGAGCTCACC 1620
               |||
Oy 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrGlyArg 294
               |||
Db 1621 AAGCTGAAGAGACCTACATTGACCCCTGCGGACCTCATACCCCGAGAGCGGCGCC 1680
               |||
Oy 295 IleAspTyrGlnSerValGlnThr-----ArgLeuIleSerAspPro 309
               |||
Db 1681 CTCACACCCCGCTTACACCGAGCGCCACGCGGAGGAGCTAGTACCTCCGATCCG 1740
               |||
Oy 310 AlaLysGlyLeuAspArgAspArgLysProIleAlaGlyArgLeuGlnAspAla 329
               |||
Db 1741 -----AACCTCCAGAACATCCCGCTCCGACCCCGCTGGGACAGAGATCCCGGGCC 1794
               |||
Oy 330 LeuLysGlnProGluLysSerValTyrLeuValSerProTyr----- 343
               |||
Db 1795 TTCATCGCCGAGAGAGGGGTGCTATTGGTGGCTGAGCTATAGCCAGATAGAGCTCAGG 1854
               |||
Oy 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspLysIleAsp 363
               |||
Db 1855 GTGGGGCCCGACCTCTCGGGCGAGCAACCTGATCCGGCTTCCAGAGAGGGGGGGC 1914
               |||
Oy 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
               |||
Db 1915 ATCCACAGGAGAACCGCCAGCTGATGTTCGGCGTCCCGGGAGGCGCTGAGAC----- 1968
               |||
Oy 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
               |||
Db 1969 -----GlyLeuLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406
               |||
Oy 394 -----GlyLeuLysLeuTyrGlnLeuGlnProAsnHisAlaVal 406
               |||
Db 1996 ATCAACTCGGGGCTCTACGGCATGTGCGCCACCGCTCTCCAGAGAGCTAGCCATC 2055
               |||
Oy 407 Pro----- 407
               |||
Db 2056 CCTTACGAGAGAGCCCGCTTCATAGAGCTTTCAGAGCTTCCCAAGGTGGG 2115
               |||
Oy 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
               |||
Db 2116 GCCTGATTGAGAGACCTGAGAGGAGGAGGCGGGGTAGCTGAGACCTCTTC 2175
               |||
Oy 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
               |||
Db 2176 -----GCGCGCGCGCGCTACGCTCCACAGCTTACAGAGCCCGGGGTG 2214
               |||
Oy 444 ArgSerAlaArgLeuAsnThrGluMetGlyValIleGlnSerPro-----Lys 460
               |||
Db 2215 AAGAGGTGCGGAGGCGGCGGAGCGC---ATGGCCTTACACATGCGCTCCGGGGCACC 2271
               |||
Oy 461 IleAlaGluGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477
               |||
Db 2272 GCGCGGACCTCATGAGCTGCTGTAAGCTTCTCCAGAGCTGAGGAAATGGGG 2331
               |||
Oy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLys 497
               |||
Db 2332 GCCAGATCTCTTCAGGTCCACGAGCGAGCTGCTCCGAGGCCCAAGAGAGGGCG 2391
               |||
Oy 498 ThrTyrProAsnGluProGluAlaLysLeuTyrLysArgIleAlaAlaLysIleLeuSer 517
               |||
Db 2392 -----GAGGCGCTGGCGGCTGCGCCAGAGAGTCTAGAGGGGGGTATCC 2439
               |||
Oy 518 Leu---LeuProIleGlu 522
               |||

```

```

Db 2440 CTGGCGGTGCCCCGTGAG 2457
RESULT 23
US-10-074-328-21
; Sequence 21, Application US/10074328
; Publication No. US20030013098A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
;              GROTELUESCHEN HALL, JEFF S.
;              LYAMICHEV, VICTOR
;              OLIVE, DAVID M.
;              PRUDENT, JAMES R.
; TITLE OF INVENTION: DETECTION OF NUCLEIC ACID SEQUENCES BY
;              INVADER-DIRECTED CLEAVAGE
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/074,328
; FILING DATE: 12-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,491
; FILING DATE: 23-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: P-40,027
; REFERENCE/DOCKET NUMBER: FORS-01802
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2502 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-074-328-21

Alignment Scores:
Pred. No.: 0.0486 Length: 2502
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
Gaps: 30

US-10-066-551-4 (1-525) x US-10-074-328-21 (1-2502)
Oy 118 AsnLeuValTyrLeuAlaGluArgGlyValArgVal----- 130
               |||
Db 961 GATCTCTGTGCGCTGCGCGCCGACGAGGGGGGCGGGGTCCACCGGGCCCCAGCCTTAT 1020
               |||
Oy 131 ArgLeuLeuLeuAspAsnAsnThrArgGlyLeu-----AspAspLeuLeu 147
               |||
Db 1021 AAAGCCTCAGGAGCTGAAGAGCGCGGGGCTTCTCGCCAAAGACCTGAGCCTTCTG 1080
               |||
Oy 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
               |||
Db 1081 GCCCTGAGGAGAACGCTTGGCTCCGCGCGGCGAGACCCCATGCTC----- 1128
               |||

```









```

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/350,597
FILING DATE: 09-Jul-1999
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-Mar-1997
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-Jan-1997
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-Dec-1996
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-Dec-1996
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-Nov-1996
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-Jul-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-Jan-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRADEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
DESCRIPTION: /desc = "DNA"
NAME/KEY: CDS
LOCATION: 1..2499
SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-10-033-297-68
Alignment Scores:
Pred. No.: 0.0486 Length: 2505
Score: 106.00 Matches: 119
Percent Similarity: 33.70% Conserves: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.91% Indels: 188
Gaps: 30
US-10-066-551-4 (1-525) x US-10-033-297-68 (1-2505)
QY 118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgVal----- 130
Db 961 GATCTTGTGCGCTGCGCCGCCAGAGGGGGCGGCTCCACCGGGCCCGAGCTTAT 1020
QY 131 ArgLeuLeuLeuAspAspAsnAsnThrArgGlyLeu-----AspAspLeuLeuLeu 147
Db 1021 AAAAGCCCTCAGGAGGAGGAGGCGGGGGCTTCTGCCAAGAACCTGAGCGCTTTC 1080
QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCGCTGAGGAGGAGGCTTGGCTCCCGCGCGAGACCCCATGCTC----- 1128
QY 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTCGCTTACCTCTGAGACCTTCCACACACACCCCGAG 1167
QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGCGCGGCTACGCGCGGAGAGTGAGAGGAGAGGAGGAGGCGCGCGCTT 1227
QY 198 -----LeuGlyArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGGCTTTCGCCAACCTGTGGGGAGGCTTGGAGGAGAGAGGCTCTTTGG 1287

```

```

QY 207 ---TyrPheLysValGlyGlu-----AspThrValPheAlaAspLeuAspIleLeuAla 223
Db 1288 CTTTACCGGAGGTGAGAGGCGCCCTTCCCGCTGTCTGCGCCACATGAG-----GCC 1341
QY 224 ThrGly-----SerValGlyGly 230
Db 1342 ACCGGGGTCCGCGCTGAGCTGGCCCTATCTCAGGGCTTGTCCCTGAGGTGGCGGGAG 1401
QY 231 ValSer-----HisAspPheAsnArgTyrTrp 239
Db 1402 ATCGCCCGCTCGAGCGCGAGGTCTTCCGCTGGCGCGGACCCCTTCAC----- 1452
QY 240 AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyGly 259
Db 1453 ---CTCACTCCCGGAGACAGCTGGAAGGCTCTTTGAC-----GAGCTAAGG 1500
QY 260 LeuGlnAlaLeuGlyTyrAsnAspGlyThrSerArgHisAla----- 273
Db 1501 CTTCGCCCATCGGCAAGACGAGAGACCGGCAACGCTCCACGCGCGCGCTCTG 1560
QY 274 -----LeuLeuArgTyrArgGlyThrVal 281
Db 1561 GAGGCCCTCCGAGGCGCCACCCCATCTGAGAGACATCTGACGTACCGGAGCTCAC 1620
QY 282 GluGln-----SerProLeuTyrGlnLysIle-----GlnThrArg 294
Db 1621 AACCTGAGAGACCTACATATTGACCTTGGCGGAGCTCATCCACCCGAGAGGCGCG 1680
QY 295 IleAspTrpGlnSerValGlnThr-----ArgLeuIleSerAspAspPro 309
Db 1681 CTCACACCGCGCTTCAACAGAGGCGGACGCGGACGCGGAGGCTGATGCTCGATCC 1740
QY 310 AlaLysGlyLeuAspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAla 329
Db 1741 -----AACCTCAGAACATCCCGCTCCGACCGCGCTTGGGACAGATCCCGCGGCC 1794
QY 330 LeuLysGlnProGlnLysSerValTyrLeuValSerProTyr----- 343
Db 1795 TTCATCGCGGAGAGGGGTGCTATGTTGGCTGCTGCTGATGACCATAGATAGCTCAG 1854
QY 344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363
Db 1855 GTGCTGGCGCCACCTCTCCGCGGAGAGAACCTGATCGGCTCTTCCAGAGAGGCGGAG 1914
QY 364 ValThrValLeuThrAsnSer-----LeuGlnAlaThrAspValAla 377
Db 1915 ATCCACAGGAGACCGCCAGCTGATGTCGGGCTCCCGGGAGGCGCTGAGC----- 1968
QY 378 AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGCGCGGGGCGGCAAGACC 1995
QY 394 -----GlyIleLysLeuTyrGlyLeuGlnProAsnHisAlaVal 406
Db 1996 ATCAACTTGGGGTCTCTTACGCGATGCGGCCACCGCTCTCCAGAGAGTACGCATC 2055
QY 407 Pro----- 407
Db 2056 CCTTACGAGAGGCGCCAGGCTTATGAGCGCTTACTTCAGAGCTTCCCAAGGTGCGG 2115
QY 408 -----AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis 423
Db 2116 GCCTGATTTGAGAGACCTTGAGAGGAGGAGGCGGGGAGTACGTGAGACCCCTTTC 2175
QY 424 AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro 443
Db 2176 -----GCGCGCGCGCTTACGTCCAGACCTTAGAGGCCCGGCTG 2214
QY 444 ArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerPro-----Lys 460
Db 2215 AAGAGGCTGGGAGGAGGCGGCGGAGCGC---ATGCGCTTCAACATGCGCTTCCAGGAG 2271
QY 461 IleAlaGlnGlnMetGluArgThrLeuAlaAspThrSerPro-----GluTyrAla 477

```

```

Db      2272 GCGCCGACATCATGAGAGCTGGCATGTGAAGCTCTTCCCACGCGCTGAGAGAATTGGGG 2331
QY      478 TyrArgValItrLeuAspAraIGHISAsnArqLauGLInTrPhisAspProLaTharIrgIs 497
Db      2332 GCCAGAGATCCTTCACGTGCAGTGCCAGAGACTGCTCTCGAGGCCCAAAAGAGAGGGCG 2391
QY      498 ThrTyProlAnglInProGLuaAlaLyLSLeuTrPLysArgILealAlaLyStleUser 517
Db      2392 -----GAGCGCCGTGGCCCCGCGCTGCCAACAGAGGTCTATGAGGGGGGTATACC 2433
QY      518 Leu---LeuProlleGIu 522
Db      2440 CTGGCCCTGCCCCCTGAG 2457

RESULT 27
US-09-940-244-68
; Sequence 68, Application US/09940244
; Publication No. US20030044796A1
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
; LENGTH: 2505
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2499)
; OTHER INFORMATION:
US-09-940-244-68

Alignment Scores:
Pred. No.:          0.0486           Length:         2505
Score:              106.00           Matches:        119
Percent Similarity: 33.70%            Conservative:    65
Best Local Similarity: 21.79%          Mismatches:     174
Query Match:        3.91%             Indels:         188
DB:                  9                Gaps:          30
US-10-066-551-4 (1-525) x US-09-940-244-68 (1-2505)

QY      118 AsnLeuValITrLeuAlaIaGLuarGLyValaIrgVal----- 130
       ::::::::::: ||||| ||||| ||||| |||||
Db      961 GATCTTTCTGGCGCCCTGGCGCGCGGAGGGGGGGCGGTCCACCGGGCCCCGAGGCTTAT 1020
       ::::::::::: ||||| ||||| ||||| |||||
QY      131 ArgLeuLeuAspAspAsnAsnTrfArgGLyLeu-----AspAspleuLeu 147
       :::: ||| ||||| ||||| ||||| |||||
Db      1021 AAAGCCCTCAGGAGACCTGAAGAGAGCGCGGGGCTTCTGCCAAAGACTGTAGCGCTTCTG 1080
       :::: ||| ||||| ||||| ||||| |||||
QY      148 AlaLeuAspSer-----HisProAsnIlleGLuValaIrgLeuPheAsnPProHe 163
       ||||| ||||| ||||| ||||| |||||
Db      1081 GCCTTAGAGGAGAGGCTCTTGCTCCGCGCGGAGAGACCCCATGCTC----- 1128
       ||||| ||||| ||||| ||||| |||||
QY      164 ValLeuArGLyStrPaIGAlaLeuGLITyrLeuThrAsp-----PheProArg 179
       ||||| ||||| ||||| ||||| |||||
Db      1129 -----CTCGCTTACTCTCTGTGACCTTTCCAACACCAACCCCGAG 1167
       ||||| ||||| ||||| ||||| |||||
QY      180 LeuAsnArGARgMeTHISAsnLySerPheThr-----AlaAspAsnArGaIaThrIle 197
       ||| :::: ||||| ||||| ||||| |||||
Db      1168 GGGGTGGCGCGCGCTCACGGCGGGGAGATGTACGAGAGAGACGGCGGAGCGGCCCTCT 1227
       ||| :::: ||||| ||||| ||||| |||||
QY      198 -----LeuGLyGLyArqAsnIlleGLyAspGLu----- 206

```

D	1228	TC	CGAGAGCTCTTCCGCAACCTGTGGGGAGGCGTTGAGGGGAGAGAGAGCTCCTTGG	128
O	207	---	TyrPheValGlyGlu-----AspThrValPheAlaIleAspIleLeuAla	223
D	1288	CTT	TACCGGAGGTGGAGAGGCGCTTTCGCGTCTCGGCCACATGTGAG-----GCC	134
O	224	ThrGly-----	serValGlyGly	230
D	1342	ACGGGGTGCGCCCTGGACGTGGCTATCTCAGGGCTTGCTCGCTGGAGGGCGGGGAG	140	
O	231	ValSer-----	HisAspPheAspArgTyrTrp	239
D	1402	ATCGCCCGCCTCGAGCGCGAGGTCTTCGCGCTGGCGGGCACCGCTTCAAC-----	145	
O	240	AlaSerHisSerAlaHisAsnAlaThrArgIleIleArgserGlysnIleGlyGly	259	
D	1453	---CTCACTCTCCGGGACACCTGGAAAGGGCTCTTTGAC-----GAGCTAGGG	1500	
O	260	LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAla-----	273	
D	1501	CTTCCCGGCATCGGCAGAGCGAATAACCGGACAGCGCTCCACAGCGCGCGCTC	156	
O	274	-----	LeuLeuArgTyrArgGluThrVal	281
D	1561	GAGCGCTCCGCGAGGCCACCGCCATGTGAGAGAGATCTCTGACATGACGGGAGCTCAC	1620	
O	282	GluGln-----	SerProLeuTyrGlnIleGly-----GlnThrArgTyr	294
D	1621	AAAGCGAAGACCTACTCATTTGACCCCTTCCGGACTCATTCACCCCGAGAGCGGCC	1681	
O	295	IleAspTrpGlnSerValGlnThr-----	ArgLeuIleSerAspAspPro	309
D	1681	CTCCACACCCGCTTCAACACAGACGGCCACGCGCCAGCGGCGCTAGTGTCTCGCATCC	1744	
O	310	AlaLysGlyLeuAspArgAspArgTyrGlyProProIleAlaIleArgLeuGlnAspAla	329	
D	1741	-----AACCTCCAGAACATCCCGCTCCGACCCCGCTTGGCGAGAGATCCGCGGCC	179	
O	330	LeuLysGlnProGluLysSerValTyrLeuValSerProTyr-----	343	
D	1795	TTTCATCGCGCGAGGGGTGGTCTATTGTGGCTTGCCCTATAGCCAGATAGAGCTCAG	185	
O	344	PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuAlaGlnAspGlyIleAsp	363	
D	1855	GTCGCGCCACCTCTCCGGGACGAGAACCTATCCGGGTCTTCCAGAGAGGGCGGCG	191	
O	364	ValThrValLeuThrAsnSer-----	LeuGlnAlaThrAspValAla	377
D	1915	ATCCACACGAGACCGCGCAGCTGATGTTGGGGTCCCGCGGAGGCGCGTGAG-----	196	
O	378	AlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAla-----	393	
D	1969	-----	CCCTGATGTGGCGCGGGCGGCAGAC	199
O	394	-----	GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaVal	406
D	1996	ATCAACTTGGGGTCTCTACGCGATGTGGCCCGCACCGCTCTCCAGAGAGCTAGCCATC	205	
O	407	Pro-----	407	
D	2056	CTTACGAGAGGCGCCAGGCGCTTATGACGCTACTTACAGACTTCCCAAGGTGCGG	2111	
O	408	-----	AlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHis	423
D	2116	GCCTGATTTGAGAGACCTCGAGAGAGGAGAGAGCGGGGATCTGAGACCCCTTTC	2176	
O	424	AlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAspPro	443	
D	2176	-----	GGCGCGCGCGCTAGCTGCGACAGCTTAAGGCGCGGGG	221
O	444	ArgSerAlaArgLeuAsnThrGluMetGlyValAlaIleGluSerPro-----Lys	460	

Db 2215 AAGAGCGTGGGAGCGCGCGAGCGC--ATGGCTTCACATGCCCGTCACAGGCGAC 2271  
Qy 461 ILeaIaGlInMeGlUaRgThrLeuAlaAspThrSerPro-----GIuTyrAla 477  
Db 2272 GCCGCGGACCCTCAATGACCTATGCTGAAGCTTCCCGAGCTGGAGAAATGGGG 2331  
Qy 478 TyrArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLys 497  
Db 2332 GCCAGGATGCTCTCAAGTCCAGCAGCAGCTGCTCTGAGAGCCCAAAAGAGAGGCG 2391  
Qy 498 ThrTyrProAsnGlnProGluAlaLysLeuTrpLysArgIleAlaLysIleLeuSer 517  
Db 2392 -----GAGCGCGTGGCGCGCTGGCCAGAGAGATGAGAGGGGTGTATCCC 2439  
Qy 518 Leu---LeuProIleGlu 522  
Db 2440 CTGGCGCTGCCCTGGAG 2457  
RESULT 28  
US-09-070-927A-150  
Sequence 150, Application US/09070927A  
Patent No. US20020120116A1  
GENERAL INFORMATION:  
APPLICANT: Charles A. Kunsch  
Patrick J. Dillon  
Steven Barash  
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 982  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/070, 927A  
FILING DATE: 04-May-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/046, 655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: 60/044, 031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: 60/066, 009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover  
REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB369  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 150:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11739 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 150:  
US-09-070-927A-150  
Alignment Scores:  
Pred. No.: 0.58 Length: 11739  
Score: 106.00 Matches: 95  
Percent Similarity: 31.12% Conservative: 55  
Best Local Similarity: 19.71% Mismatches: 162  
Query Match: 3.91% Indels: 170

DB: 10 Gaps: 22  
US-10-066-551-4 (1-525) x US-09-070-927A-150 (1-11739)  
Qy 42 GIUGlUaRgThrGlnSerArgHisPheAsnTrpSerLysProValArgLeuAspAsnIle 61  
Db 3963 GAAGCGCTTAACACAAAGTCTTAAACGATGGGTCAAAACCCCTT----- 4007  
Qy 62 LeuGlnIleArgHisThrProHisThrAsnGlnLeu-----SerAspIleTyrLeuLeu 79  
Db 4008 -----ACGAGATGATCCAGTAACATTTATCGCAAT 4040  
Qy 80 AsnAspProHisGluAlaPheAlaIaArgAlaAlaLeuIleGluSerAlaGlnHisSer 99  
Db 4041 CCAGAAACGATGAGATTTCCGACCTTACACTGGC-----GATATTTGAAGAAATGAC 4094  
Qy 100 LeuAspLeuGlnTyrTyrIleTrp-----ArgAsnAspIleSerGlyArgLeuLeu 116  
Db 4095 GTCAAAAAGCTCAAGCTGAATGAGAGAAAGCCCAAGCGATGTGTAAAAAGTGAAA 4154  
Qy 117 PheAsnLeuValTyrLeuAlaIaLaGluArgGlyValArgValArgLeuLeuAspAsp 136  
Db 4155 CTTTCATTGCTCGCGCGACACACAGATCAAGAAACGATTCGATATGATTCAGAACT 4214  
Qy 137 AsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGlu 156  
Db 4215 CAGTTGCAAGAAATCTCGCAGGTTTAGAAATTTACATTTCCGACACAGATATAT 4274  
Qy 157 ValArgLeuPheAsnProPheValLeuArgLysTTPArgAlaLeuGlyTyrLeuThrAsp 176  
Db 4275 GCG----- 4277  
Qy 177 PheProArgLeuAsnArgArgmethHisAsnLysSerPheThrAlaAspAsnArgAlaThr 196  
Db 4278 -----AACCAATCGCGACGCTGAATAATTAAGATTTGCTCTTCAGAGATGG 4325  
Qy 197 IleLeuGlyGlyArgAsnIleGlyAspGlyTyrPheLysValGlyLysAspThrValPhe 216  
Db 4326 ATTGCCGCGAGTAGAATTA---GACTCTTACTTAACCTA----- 4364  
Qy 217 AlaAspLeuAspIleLeuAlaThrGlySerValValGlyGlyValSerHisAspPheAsp 236  
Db 4365 -----TATGCAGAGAAATCAAGTTACATTAACGCGC 4394  
Qy 237 ArgTyrTrpAlaSerHisSerAlaHisAsnAla-----Thr 248  
Db 4395 AATTAT-----CATATGCAAAATACGACCAATTTGTTAGAGAGCA 4436  
Qy 249 ArgIleIleArgSerGlyAsnIleGlyGlyLeuGlnAlaLeuGlyTyrAsnAspGlu 268  
Db 4437 CGAAGATTAATGCCAATATATCCAGAGAAA-----CAGTTGCGAGAAATCAAGAGCG 4490  
Qy 269 ThrSerArgHisAlaLeuLeuArgTyrArgGlyIuThrValGluGlnSerProLeuTyrGln 288  
Db 4491 GAAGAC-----ATCTTGTAACCAAGATGCTGCCCAAGTACCGCTGAT--- 4535  
Qy 289 LysIleGlnThrGlyArgIleAspTrpArgLysSerValGlnThrArgLeuIleSerAsp 308  
Db 4536 -----CAAAAGTGCCTCAAAATATCTATCAAT 4562  
Qy 309 ProAlaLysGlyLeuAspArgAspArgArgLysProIleAlaGly-----LysGln 332  
Db 4563 -----CCTAAATTAAGGCAATTGATTATCAG 4589  
Qy 325 -----ArgLeuGlnAspAlaLeu-----LysGln 332  
Db 4590 TTGTATGGGAGATTATTTCCACTTTCGCAATGCTTATTAACAGAAATGATTCAGAAAG 4649  
Qy 333 ProGluLysSerValTyrLeuValSerProTyrPhe----- 344  
Db 4650 CAGTAAAAAATGAAAAAATTAATTCGACAGCATTTTAACCTTACGTAAGAAAGTACCCGCT 4709  
Qy 345 -----ValProThrLysSerGlyThrAspAlaLeuAlaLysLeu 357

```

Db 4710 TCAGATGCCACAGTCAACACACACAGGAGCGGTGGGCAAGTGGTTTACCAAAATG 4769
QY 358 ValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAla 377
Db 4770 ATTGAA-----ACAGAACTTCATGATATAGCTTGGCC 4802
QY 378 AlaValHis-----SerGlyTyrValLysTyrArgLysProLeuLeuLysAla 393
Db 4803 GATGTGCATTATTAACGAAACAAACGGCTTTTAAACGGACGGTGGCT----- 4850
QY 394 GlyIleLysLeuTyrGlnLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly 413
Db 4851 -----GTAACCAACCTGCGACGAGGATATCGGC 4880
QY 414 LeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleValAspLysArg 433
Db 4881 TTG-----ATTGCACATCTGGATACGGGGATTTTTCAGCAGAAACATCCGT 4928
QY 434 IlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMetGly 453
Db 4929 CCCCAGTCACTACT---AATTATGATGTGTCAAGATGTCTCTTGAATCAACAGAACAGGC 4985
QY 454 ValValIle-----GlnSerProLysIleAlaGlnGlnMetGlnArgThrLeuAla 470
Db 4986 ATCGTTTATTCAGTGTGATTTTCCCAATTTAAAGAGTATCAAGGGAAACGTGATT 5045
QY 471 AspThr 472
Db 5046 ACTACA 5051

```

RESULT 29  
 US-10-033-297-65  
 Sequence 65, Application US/10033297  
 Publication No. US20020187486A1  
 GENERAL INFORMATION:  
 APPLICANT: Hall, Jeff G.  
 Lyamichev, Victor I.  
 Mast, Andrea L.  
 Brow, Mary Ann D.  
 TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple  
 Sequential Invasive Cleavages  
 NUMBER OF SEQUENCES: 163  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Medgen & Carroll, LLP  
 STREET: 220 Montgomery Street, Suite 2200  
 CITY: San Francisco  
 STATE: California  
 COUNTRY: United States Of America  
 ZIP: 94104  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/100/033, 297  
 FILING DATE: 12-NOV-1996  
 CLASSIFICATION: <unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/09/350,597  
 FILING DATE: 09-JUL-1999  
 APPLICATION NUMBER: US/08/823,516  
 FILING DATE: 24-MAR-1997  
 APPLICATION NUMBER: PCT/US97/01072  
 FILING DATE: 21-JAN-1997  
 APPLICATION NUMBER: US 08/759,038  
 FILING DATE: 02-DEC-1996  
 APPLICATION NUMBER: US 08/758,314  
 FILING DATE: 02-DEC-1996  
 APPLICATION NUMBER: US 08/756,386  
 FILING DATE: 29-NOV-1996  
 APPLICATION NUMBER: US 08/682,853

```

FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 65:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2499
SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-10-033-297-65
Alignment Scores:
Pred. No.: 0.0639 Length: 2505
Score: 105.00 Matches: 116
Percent Similarity: 33.52% Conservative: 65
Best Local Similarity: 21.48% Mismatches: 183
Query Match: 3.87% Indels: 176
Gaps: 27
US-10-066-551-4 (1-525) x US-10-033-297-65 (1-2505)
QY 118 AsnLeuValTyrLeuAlaAlaGlnArgLysValArgVal----- 130
Db 961 GATCTTCTGGCCCTGCGCCGCGCCGAGGGGGGGGTCCACCGGGCCCGAGCGCTTAT 1020
QY 131 ArgLeuLeuLeuAspAspAsnAsnThrArgLysLeu-----AspAspLeuLeuLeu 147
Db 1021 AAGGCTCTCAGGAGACCTGAGAGAGCGCGGGCTTCTCGCCAAAGACCTGAGGTTCTG 1080
QY 148 AlaLeuAspSer-----HisProAsnIleGluValArgLeuPheAsnProPhe 163
Db 1081 GCCCTGAGGAGAGCGCTTGGCTCCCGCCGCGAGCGAGCCCATGCTC----- 1128
QY 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp-----PheProArg 179
Db 1129 -----CTCGCTTACTCTCGGACCTTCCACACACACCGCCGAG 1167
QY 180 LeuAsnArgAlaGlnMetHisAsnLysSerPheThr-----AlaAspAsnArgAlaThrIle 197
Db 1168 GGGGTGGCCCGCGCTCAGCGGGAGAGTGCAGAGAGGCGGGGAGACCGCGCGCTT 1227
QY 198 -----LeuGlyLysArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGCGCTTCCGCAACCTGTGGGGAGGCTTGAAGGGGAGAGAGCGCTCTTGG 1287
QY 207 --TyrPheLysValGlyGlu-----AspThrValPhe 216
Db 1288 CTTTACCGGAGGTGAGAGAGCGCCCTTCCGCTGCTGCGCCACATGAGACCGACGGG 1347
QY 217 AlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAspPheAsp 236
Db 1348 GTGGCCCTGAGCGTGGCTATATCAGGCGCTTGTCCCTGAGGTGCGCGAGAGATCGCC 1407
QY 237 ArgTyrTrpAlaSer-----HisSerAlaHis 245
Db 1408 CGCCTCGAGCGCGAGGCTTCCGCTGCGCGCCAGCCACCTTCAACCTCACTCCCGGAC 1467
QY 246 AsnAlaThrArgLysIleIleArgSerGlyAsnIleGlyLysGlyLeuGlnAlaLeuGlyTyr 265
Db 1468 CAGGTGAAGAGGTCTCTTGGAC-----GAGTAGGCGCTTCCCGCATGCGCAG 1518

```

```

QY 266 AsnaSpGluThrSerArgHisAla----- 273
DB 1519 ACGGAGAGACCGGACGCTCCAGACGCCGCCGCTCGAGAGCCCTCCGCGAGGCC 1578
QY 274 -----LeuLeuArgTyrArgGluThrValGluIn----- 283
DB 1579 CACCCATCGTGAGAGATCTCGACGTACCGGAGCTCAGCAGAGCTGAAGACACTTAC 1638
QY 284 ---SerProLeuTyrGlnLysIle-----GlnThrGlyArgIleAspIleAspIleAspIle 300
DB 1639 ATTGACCCCTTCCGCGACCTCATCCACCCAGAGAGCGGCCCTCCACACCCGCTTCAAC 1698
QY 301 GlnThr-----ArgLeuIleSerAspAspProAlaThrArgLysThrTyrProAsnGluPro 315
DB 1699 CAGACGGCCACGCGCCAGCGGCTAGTACCTCCGATCC-----AACCTCCAGAAC 1752
QY 316 AspaTgaTgTysProProIleAlaGluArgLeuGlnAspAlaLeuLysGlnProGluLys 335
DB 1753 ATCCCGCTCCGACCCCGCTTGGCAGAGATCCCGCGGCTTCATCCCGCAGAGAGGG 1812
QY 336 SerValTyrLeuValSerProTyr-----PheValProThrLysSer 349
DB 1813 TGGCTATTGGTGGCCCTGGACCTATAGCCAGATAGACTCAGGGTGCTGGCCCACTCTCC 1872
QY 350 GlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsn 369
DB 1873 GCGCAGACGAGAACCTATCCGCTTCCAGAGAGCGGCGGACATCCACAGCGAGACCGCC 1932
QY 370 Ser-----LeuGlnAlaThrAspValAlaAlaValHisSerGlyTyr 383
DB 1933 AGCTGGATGTTGCGGCTCCCGCGGAGCGCGTGAC----- 1968
QY 384 ValLysTyrArgLysProLeuLeuLysAla----- 393
DB 1969 -----CCCGTATGGCGCGCGCGCGCGCCAGACATCACTTCGGGGTCTTC 2013
QY 394 ---GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValPro----- 407
DB 2014 TACGGCATGTGGCCCGCCACCGCTTCCCGAGAGCTAGCATCCCTACGAGAGGCCAG 2073
QY 408 ----- 409
DB 2074 GCCTTCATTGACCGCTACTTTCAGAGCTTCCCAAGGTGCGGCGCTGATTGAGAGACC 2133
QY 410 LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleVal 429
DB 2134 CTGGAGAGAGGCGAGAGCGGGGCTAGCTGAGACCTCTTC----- 2175
QY 430 AspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsn 449
DB 2176 ---GGCGCGCGCGCTAGCTGACAGACCTAAGAGCCCGGCTGAAGAGCGTGGAGAGGG 2232
QY 450 ThrGluMetGlyValValIleGluSerPro-----LysIleAlaGluGlnMetGlu 466
DB 2233 GCCGAGCGC---ATGGCTTCAACATGCCGTCAGAGCGACCGCCGACCATCATGAG 2289
QY 467 ArgThrLeuAlaAspThrSerPro-----GluTyrAlaTyrArgValThrLeuAsp 483
DB 2290 CTGGCTATGTGTAAGCTTCTCCAGGCTGAGAGAAATGGGGCGAGATGCTCTTCAG 2449
QY 484 ArgHisAsnArgLeuGlnThrPheAspProAlaThrArgLysThrTyrProAsnGluPro 503
DB 2350 GTCCACAAACAGAGCTGTGCTCGAGCGCCCAAAAGAGAGGGCG-----GAGGCC 2397
QY 504 GluAlaLysLeuThrLysArgIleAlaAlaLysIleLeuSerLeu---LeuProIleGlu 522
DB 2398 GTGGCGCGCGCTGCGCAAGAGGTCAATGAGAGGGGTGTATCCCTGCCGCGCCCTCGAG 2457

```

```

RESULT 30
US-10-033-297-70
; Sequence 70, Application US/10033297
; Publication No. US20020187486A1
; GENERAL INFORMATION:

```

```

APPLICANT: Hall, Jeff G.
            Lyamichev, Victor I.
            Mast, Andrea L.
            Brow, Mary Ann D.
TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
                    Sequential Invasive Cleavages
NUMBER OF SEQUENCES: 163
CORRESPONDENCE ADDRESS:
ADDRESS: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/033,297
FILING DATE: 12-NOV-1999
CLASSIFICATION: <Unknown>
PRIORITY DATA:
APPLICATION NUMBER: US/09/350,597
FILING DATE: 09-JUL-1999
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-MAR-1997
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 2505 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2499
SEQUENCE DESCRIPTION: SEQ ID NO: 70:
US-10-033-297-70
Alignment Scores:
Pred. No.: 0.0639 length: 2505
Score: 105.00 matches: 119
Percent Similarity: 33.708 conservative: 65
Best Local Similarity: 21.798 mismatches: 174
Query Match: 3.87% indels: 188
DB: 9 gaps: 30
US-10-066-551-4 (1-525) x US-10-033-297-70 (1-2505)
QY 118 AsnLeuValTyrLeuAlaGluArgGlyValArgVal----- 130

```





```

QY 118 AsnLeuValTyrLeuAlaIaIaGluArgGlyValArgVal----- 130
   ::::: ||||| ||||| |||||
Db 961 GATCTTCTGCGCCCTGCGCCCGCCAGGGGGGGGTCACCGGCCCCCGAGCTTAT 1020
QY 131 ArgLeuLeuLeuAspAsnAsnThrArgGlyLeu----- AspAspLeuLeu 147
   ::::: ||||| ||||| |||||
Db 1021 AAAAGCCCTCAGGACCTGAAGAGGGGGGCTTCTGCCAAAGACTGACGCTTCTG 1080
QY 148 AlaLeuAspSer----- HisProAsnIleGluValArgLeuLeuAsnProPhe 163
   ||||| ||||| |||||
Db 1081 GCCCTGAGGAAAGCCTTGCCCTCCCGCGGAGAGCCCAATGCTC----- 1128
QY 164 ValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp----- PheProArg 179
   ||||| ||||| |||||
Db 1129 -----CTCGCTACCTCTCTGACCTTCCACACACACACCCCGAG 1167
QY 180 LeuAsnArgArgMetHisAsnLysSerPheThr----- AlaAspAsnArgAlaThrIle 197
   ||||| ||||| |||||
Db 1168 GGGGTGCGCGCGCTACGCGGGAGTGAGGAGAGGAGGAGGAGGCGCGCCCTT 1227
QY 198 -----LeuGlyArgAsnIleGlyAspGlu----- 206
Db 1228 TCCGAGAGGCTCTTGCCCAACCTGTGGGGAGGCTTGAGAGGAGAGAGGCTCTTGG 1287
QY 207 ---TyrPheLysValGlyGlu-----AspThrValPhe 216
   ||||| ||||| |||||
Db 1288 CTTTACCGGAGGTGAGAGGCGCCCTTCCGCTGTCTGCGCCACATGAGGCGCAGGGG 1347
QY 217 AlaAspLeuAspIleLeuAlaThrGlySerValValGlyLysSerHisAspPheAsp 236
   ||||| ||||| |||||
Db 1348 GTGCGCTGAGAGGCTCTATCTCAGGGCTTGTCCCTGAGGTGCGCAGGAGATCGCC 1407
QY 237 ArgTyrTrpAlaSer-----HisSerAlaHis 245
   ||||| ||||| |||||
Db 1408 CGCTTCGAGGCGAGTCTTCCGCGCGGCGCCACCTCAACTCACTCCGCGAGC 1467
QY 246 AsnAlaThrArgIleIleArgSerGlyAsnIleGlyLysGlyLeuAlaLeuGlyTyr 265
   ||||| ||||| |||||
Db 1468 CAGCTGGAAGGGTCTCTTGAC-----GAGCTAGGGGTCTCCGCGCATCGGCAG 1518
QY 266 AsnAspGluThrSerArgHisAla----- 273
   ||||| ||||| |||||
Db 1519 ACGGAGAAAGACCGGAGGCTCACACGCGCGCTCTGAGAGGCTCCGCGAGGCC 1578
QY 274 -----LeuLeuArgTyrArgGluThrValGluGln----- 283
   ||||| ||||| |||||
Db 1579 CACCCCATCGTGAGAGATCTGAGTACCGGAGCTCACCAAGCTGAAGACACTTAC 1638
QY 284 ---SerProLeuTyrGlnLysIle-----GlnThrGlyArgIleAspTrpGlnSerVal 300
   ||||| ||||| |||||
Db 1639 ATTGACCCCTTGCGGACCTCATCCACCCAGAGGCGCGCTCCACACCCGCTTCAAC 1698
QY 301 GlnThr-----ArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315
   ||||| ||||| |||||
Db 1699 CAGACGCGCACGCGCACGCGGAGGCTAGTACCTCCGATCCC-----AACCTCCAGAAC 1752
QY 316 AspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlyLys 335
   ||||| ||||| |||||
Db 1753 ATCCCGCTCCGACCCCGCTTGCGGAGAGGATCCGCGGCTTCACTCCCGAGGAGGGG 1812
QY 336 SerValTyrLeuValSerProTyr-----PheValProThrLysSer 349
   ||||| ||||| |||||
Db 1813 TGGCTATTGTGGTGGCTGAGCTATACCAATAGAGCTCAGGCTGCTGGCGCCACTCTCC 1872
QY 350 GlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsn 369
   ||||| ||||| |||||
Db 1873 GGGGACGAGAACTGATCCGGGCTTCCAGAGGGGCGGAGACTTCCACACGAGACCGCC 1932
QY 370 Ser-----LeuGlnAlaThrAspValAlaIaIaValHisSerGlyTyr 383
   ||||| ||||| |||||
Db 1933 AGCTGATGTTCGGGCTCCCGCGGAGGCGGTGAC----- 1968

```

```

QY 384 ValLysTyrArgLysProLeuLeuLysAla----- 393
Db 1969 -----CCCTGATGCGCGGGGGCGCCAGAGCATCACTTCCGGGTCTTC 2013
QY 394 ---GlyIleLysLeuTyrGluLeuGlnProAsnHisAlaValPro----- 407
   ||||| ||||| |||||
Db 2014 TAGGATGTGCGCCACCGCTCTCCAGAGCTAGCCATCCCTTACAGAGAGGCCAG 2073
QY 408 -----AlaThr 409
Db 2074 GCCTTATGAGCGCTACTTTCAGAGCTTCCCAAGTGGCGGCTGATGAGAACCC 2133
QY 410 LysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheIleVal 429
   ||||| ||||| |||||
Db 2134 CTGAGAGAGGCGAGAGGGGGGTACTGTGAGACCTCTTC-----f----- 2175
QY 430 AspGlyLysArgIlePheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsn 449
   ||||| ||||| |||||
Db 2176 ---GGCGCGCGCGCTACGTCGCGAGCTAGAGGCGCGGTGAAAGAGCGTGGGAGGCG 2232
QY 450 ThrGluMetGlyValValIleGluSerPro-----LysIleAlaGluGlnMetGlu 466
   ||||| ||||| |||||
Db 2233 GCGGAGCGC---ATGGCTTCAACATGCCGTCAGAGGACCGCGCGCACTCATGAAG 2289
QY 467 ArgThrLeuAlaAspThrSerPro-----GluTyrAlaTyrArgValThrLeuAsp 483
   ||||| ||||| |||||
Db 2290 CTGCTATGTGTAAGCTCTTCCCGAGGCTGAGAAATGGGGCCAGATGCTCTTCAAG 2349
QY 484 ArgHisAsnArgLeuGlnTrpHisAspProAlaThrArgLysThrTyrProAsnGluPro 503
   ||||| ||||| |||||
Db 2350 GTCCACACAGAGCTGCTCTCGAGGCCCAAAAGAGAGAGCG-----GAGGCC 2397
QY 504 GluAlaLysLeuTrpLysArgIleAlaIaLysIleLeuSerLeu---LeuProIleGlu 522
   ||||| ||||| |||||
Db 2398 GTGCGCGCGCTGCGCAAGAGAGTATGAGAGGGGTATTCCTTGGCCGTGCGCTGAG 2457

RESULT 32
US-09-940-244-70
; Sequence 70, Application US/09940244
; Publication No. US20030044796A1
; GENERAL INFORMATION:
; APPLICANT: Neri, Bruce P.
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor
; APPLICANT: Smith, Lloyd M.
; TITLE OF INVENTION: Reactions on Dendrimers
; FILE REFERENCE: FORS-06478
; CURRENT APPLICATION NUMBER: US/09/940,244
; CURRENT FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 2505
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: CDS
; LOCATION: (1)..(2499)
; OTHER INFORMATION:
US-09-940-244-70

Alignment Scores:
Pred. No.: 0.0639 Length: 2505
Score: 105.00 Matches: 119
Percent Similarity: 33.70% Conservative: 65
Best Local Similarity: 21.79% Mismatches: 174
Query Match: 3.87% Indels: 188
DB: 9 Gaps: 30

US-10-066-551-4 (1-525) x US-09-940-244-70 (1-2505)

```



```

: SEQ ID NO 9187
: LENGTH: 1461
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1461)
: US-09-815-242-9187

Alignment Scores:
Pred. No.: 0.0405 Length: 1461
Score: 103.50 Matches: 77
Percent Similarity: 32.59% Conservative: 55
Best Local Similarity: 19.01% Mismatches: 114
Query Match: 3.81% Indels: 159
DB: 10 Gaps: 20

US-10-066-551-4 (1-525) x US-09-815-242-9187 (1-1461)

QY 75 AspleryleuAsnAspPro-----HisGluAlaPheAlaAlaArgAlaAla 91
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 GATTGGTTTGGTATGAAACCGATTGTGAGGTGAGAGAGACCTTACCCCATGTCAG 363

QY 92 LeuilegSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleTyrPArgAsnAspIle 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 364 TTGGTCGAAACGGCT----- 378

QY 112 SerGlyArgLeuLeuPheAsnLeuVal-----TyrLeuAlaAlaGluArgGlyVal 128
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 379 -----CTTTGAAACATCGTCACATACAGACCTTGCGCAGCAAGCACACT 426

QY 129 ArgValArgLeuLeuAsnAspAsn-----AsnThrArgGlyLeuAsp 143
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 427 CGATTTCGTTGCTTATCGAAGATGAACCTTGATGGATTGGACACGCTGGCTCAA 486

QY 144 AspleuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnProPhe 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 487 GAAATGATCGCGCCATC----- 504

QY 164 ValLeuArgLysThrPArgAlaLeuGlyTyrIleuThrAspPheProArgLeuAsnArgArg 183
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 505 -----TGG----- 507

QY 184 MethIsanLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyLysArgAsnIle 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 508 -----GGAACACGCGAGCTGTGATTTGGTGGCGCCAT--- 540

QY 204 GlyAspGluTyrPheLysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAla 223
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 GGAACCAACAACGTCGCGGCGTAAG-----CTTTTGACATTCCTGTTTGGGA 591

QY 224 ThrGlySer-----ValValGlyLysIleSerHisAspPheAspArgTyr 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 592 ACCCATGCGCCATCGCTTGGTACAGGTTTATGCAATGCAAGCTTTCAGAGCTTAC 651

QY 239 TrpAlaSerHis-----SerAlaHisAsnAlaThrArgIle--- 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 652 GCTGGAGACCAACAAATTGTGTCTTCTTGATACCTTATGACACCTTCGATCGGT 711

QY 250 ----- 250

Db 712 GTACAGCTGCCAATTCAGGTGCGCGTACGTCGATGATCAGATTAACTTATGGGTGTG 771

QY 251 ---IleArgSerGlyAsnIleGly-----LysGlyLeuGlnAla 262
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 772 CGGATTGACTGTGGGATAATTGGCTACATTCTTAAGAAAGTCCTGACCAACTGGACAG 831

QY 263 LeuGlyTyrAsnAspGluThrSerArgHisAlaLeuLeuArgGluThrValGlu 282
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 832 GCTGATTTACAGAGCT-----AAGATTATGCTTAAATGATTTGGACGAA 879

QY 283 GlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrpGlnSerValGlnThr 302
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 880 AATATCTCTCAACCTCAAGATGCAAAAGCCAAAGATTGATGCTGGCGGTGATAC 939
QY 303 ArgLeuIleSer-----AspAspProAlaLysGly-----Leu 313
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 940 AAGCTGATTACAGCTTATGACCGACCGACCTCTTGGGCGGTTTACAGATTGTTGCAATC 999

QY 314 AspArgAspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnPro 333
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1000 GAAGATGAA-----ACTGTGATGATGCCAATACGATTAACTGTCT 1041

QY 334 GluLysSerValTyrLeuValSerPro-----TyrPheValProThr 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1042 AATAATGCGGAAAAAGTGTGACAGCGTGAAGACAGATGTGGCCATTACCAAGTGT 1101

QY 348 LysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeu 367
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1102 GAAAAGGCAAGTCAGAAAGTGACTACATCTTATGATGCTGTGATATTAGC----- 1155

QY 368 ThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArg 387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1156 -----GACATACAGAAATCAAGATGTTCCATCGACCTATACATCACTCAAG 1203

QY 388 Lys-----ProLeuLeu-----LysAlaGlyIle 395
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1204 AAGACGTTCCGTAATTGATGCCGCTCTCTCTGTGATATCTTCAAGAAAGTATA 1263

QY 396 LysLeuTyrGlnLeu 400
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1264 TTAGTTTACAACTTG 1278

RESULT 34
US-09-938-842A-989
: Sequence 989, Application US/09938842A
: Patent No. US20020160378A1
: GENERAL INFORMATION:
: APPLICANT: Harper, Jeff
: APPLICANT: Kreps, Joel
: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT300-3
: CURRENT APPLICATION NUMBER: US/09/938, 842A
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227, 866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264, 647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300, 111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 989
: LENGTH: 1377
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-938-842A-989

Alignment Scores:
Pred. No.: 0.0422 Length: 1377
Score: 103.00 Matches: 101
Percent Similarity: 37.05% Conservative: 75
Best Local Similarity: 21.26% Mismatches: 172
Query Match: 3.80% Indels: 128
DB: 9 Gaps: 23

US-10-066-551-4 (1-525) x US-09-938-842A-989 (1-1377)

QY 5 ProLysThrGln-----AlaMetProSerGluThrIleSerLeuMetLysThrArgSer 22
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 4 CCTGCGCTACAAATGCTTACAAATGCGGACCTTACTAATTTCTCTCTCAACCTCTGAGT 63

QY 23 LeuIleSerLeuLeuLysLeuLeuLysSerCysSerSerTrpLeuPro----- 39
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Db 64 TTAGAGCTCACTATTCGTCCCGGCGATGATCACCCGTTTACCGGTCACCC 123  
 QY 40 -----ProleugluValThrGluSerArg 48  
 Db 124 CGATCGGGTTTACAGACTACTACAGTTATGCTACTATGACGAGAACAAACAGAGACT 183  
 QY 49 HisPhe-----AsnThrSerLysProValArgLeuAspAsnIleLeuGln 63  
 Db 184 CTGTTTACTACTGCTGGTGAAGCAGAAACCAACCCACTCTAAACCTCTGCTCTGG 243  
 QY 64 ILeaRghNstHrProHstHrAsnGlyLeuSer---AspIleTyrlLeuLeuAspPro 82  
 Db 244 CTCATGAGGACACCTGGATGCTTCATCATTAGCTGTGTGATCTCTGACAAATGAGACT 303  
 QY 83 HisGluAaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSerLeuAspLeu 102  
 Db 304 -----TTTACACCAAGAGATCAATCTTGATGAGAAATCAACACAGC----- 345  
 QY 103 GlnTyrrTyrlleTPraGAsnAspLieserGlyArgLeuLeuPheAsnLeuValTyrlleu 122  
 Db 346 -----TGAATCAAGAGCT-----AAATCTGTGATCTG 375  
 QY 123 AlaAlaGluArgGlyValArgValArgLeuLeuAspAspAsnThrArgGlyLeu 142  
 Db 376 GAGACACCTGTGGAGTTGGATTCTTATGCAATGAGACTCTCTTATGAGGGTGTG 435  
 QY 143 AspAspLeuLeuAlaLeuAspSerHisProAsnIleGluValArgLeuPheAsnPro 162  
 Db 436 AATGATTAAGTCACTGCACAAAGAC-----AATCTT 465  
 QY 163 PheValLeuArgLysTTPraGAlaLeuGlyTyrlleuThrAspPheProArg---LeuAsn 181  
 Db 466 GTGTTCTGCAAAATG-----TTCTCAAAATTCCTCAATATCTCAAC 510  
 QY 182 ArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIleLeuGlyGlyArg 201  
 Db 511 AGAGTCTCTTC-----ATCACTGTGAAGCTATGCTGGCCA-TTATGTCTC--- 557  
 QY 202 AsnIleGlyAspGlyTyrrPheLysValGlyGluAspThrValPheAla-AspLeuAspIle 221  
 Db 558 TCACCTTGCTCAGCTTATGATTGATTCATACACAGACACCAATTTGTTTATCTCAAGG 617  
 QY 221 eleAlaThrGlySerValValGlyGluValSerHisAspPheAsp-----ArgTy 238  
 Db 618 CATTCGATTGGCAATCCGGTTATGAGTTTGCAACGGATTTCATTCACGACAGAGTA 677  
 QY 238 rTPaLaseHis-----SerAlaHisAs 246  
 Db 678 TTTCGTCTCAGCGGTTGATATGCAATCAACTTACAATTTTCACATCTTCTTGA 737  
 QY 246 AlaThrArgIleIleArgSerGlyAsnIleGly-----LysG 259  
 Db 738 TTACTCGAGATTGTTGAGTAGATGATCAGATGATGCGAGTATGTCACATAAAGT 797  
 QY 259 yLeuGluAlaLeuGlyTyrrAsnAspGluThrSerArgHisAlaLeuAsnArgTyrrArgG 279  
 Db 798 GCTAAGTCAAAGTTGGCATC-----GAAACAGATGATTC---ATAGCAATATGAGCT 848  
 QY 279 uThrValGlu-----GlnSerProLeuTyrrGlnLysIleGlnThrArgIleAs 296  
 Db 849 TACCTTGACCTGTCATCCCTCTGCTGCTATCTCATCCAAACAGAGGGA----- 900  
 QY 296 pTrrGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAs 316  
 Db 901 -----GAGACGGTGGATGTG---TGTCTAAGAGATGAGCGGTGAATATATTACCGGAG 953  
 QY 316 ParArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSe 336  
 Db 954 AGATGTCAAAAGCTCTCACGCTCGCT----- 984  
 QY 336 rValTyrlleuValSerProTyrrPheValProThrLysSerGlyThrAspAlaLeuAla 356  
 Db 985 -----GTGCGAACTCGCAATGAGACTGTTTTCACGAGATGT 1019

QY 356 sleuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspVa 376  
 Db 1020 GCTGATTACAGACTACTGATGTGAGAGTACCAACGATTAATATCTAGAGACT----- 1074  
 QY 376 lAlaAlaValHisSerGlyTyrrValLysTyrrArgLysProLeuLeuLysAlaGlyIle 396  
 Db 1075 -----CTGTAAAGCTGCTGTTCC 1094  
 QY 396 s-----LeuTyrrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGly 414  
 Db 1095 AGTTTGTGTGACGTGAGATCAAGACTCTGTGATTCGTTAACCGAAGTGAACATT 1154  
 QY 414 uThrGlySerSerValThrSerLeuHisAlaLysThrPheIle 428  
 Db 1155 AGTGAAGGACTGGCTGAGAGTTGGGACCTAGAAACAACCGTG 1197  
 RESULT 35  
 US-09-883-825-50  
 Sequence 50, Application US/09883825  
 Patent No. US20020151024A1  
 GENERAL INFORMATION:  
 APPLICANT: Beavo, Joseph A.  
 Bentley, Kelley  
 Charbonneau, Harry  
 Sonnenburg, William K.  
 TITLE OF INVENTION: DNA Encoding Mammalian  
 Phosphodiesterases  
 NUMBER OF SEQUENCES: 58  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
 Bicknell  
 STREET: Two First National Plaza, 20 South Clark  
 Street  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: USA  
 ZIP: 60603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/883,825  
 FILING DATE: 18-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 09/123,783  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: 08/297,494  
 FILING DATE: <Unknown>  
 APPLICATION NUMBER: US 07/688,356  
 FILING DATE: 04-APR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: No. US20020151024A1and, Greta E.  
 REGISTRATION NUMBER: 35,302  
 REFERENCE/DOCKET NUMBER: 27866/30822  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 346-5750  
 TELEFAX: (312) 984-9740  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 50:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2693 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 176..2077  
 SEQUENCE DESCRIPTION: SEQ ID NO: 50:

US-09-883-825-50

## Alignment Scores:

Pred. No.:	0.124	Length:	2693
Score:	103.00	Matches:	119
Percent Similarity:	34.188	Conservative:	84
Best Local Similarity:	20.03%	Mismatches:	213
Query Match:	3.80%	Indels:	179
DB:	10	Gaps:	28

US-10-066-551-4 (1-525) x US-09-883-825-50 (1-2693)

```

OY      8 GlnAlaMetProSerGluThrIleSerLeuMetLysThrArgSerLeuIleSerLeu 27
DB      432 CAGAGATGAGCTCAGTACATTCATTCAGATG-----CTG 467
OY      28 CysLeuLeuLeuLeuSerCysSerSerProProLeuGluGluArgThrGlySer 47
DB      468 TCCCTCTGAGG---TCCAGAGCTGCTGCTCCACCTTCACCGCGCAGATG-GGGATG 523
OY      48 ArgHisPheAsnThrSerLysProValArgLeuAspAsnIleLeu-----62
DB      524 ATGCTCAGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 583
OY      63 -----GlnIleArgHisThrProHisThrAsnGlyLeuSer-----74
DB      584 GGGATATTGTTGAGAGAGATGTATAGACGACATCAACATGCTGGAGCTAGCTATGCA 643
OY      75 -----AspIleTyrIleLeu 79
DB      644 CCAAGCTGTTATGAGGAGATTAAGAGATGTGACAGAGTGTCTTTCAGCTTTTCCCTC 703
OY      80 AsnAspProHisGluAlaPheAlaAlaArgAlaAlaLeuIleGluSerAlaGluHisSer 99
DB      704 AATGAG-----GCCAGTGGGAGTCAATGCA 727
OY      100 LeuAspLeuGluTyrTyr-----IleTrrArgAsnAspIleSerGlyArgLeuPhe 117
DB      728 CTGAATATTATTCTGATGACTCAGACACTGATGATGATGATGATGATGATGATGATG 787
OY      118 AsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArgLeuLeuLeuAspAsn 137
DB      788 CCCATTTCGACCTGTCTCA-----808
OY      138 AsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSerHisProAsnIleGluVal 157
DB      809 -----TTTGTGAGAGCCCTGGAAGTGGATGACAGACACACAAA-----847
OY      158 ArgLeuPheAsnProPheValLeuArgLysTrrArgAlaLeuGlyTyrIleThrAspPhe 177
DB      848 -----AATCTTTAC-----856
OY      178 ProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThrIle 197
DB      857 -----CATTAAGCTTAATGACGCTGCGGATGTTACACAGACAGT 895
OY      198 LeuGlyGlyArgAsnIleGlyAspGluTyr-----PheLysValGlyGluAspThrVal 215
DB      896 -----CATTAAGCTTCTTATTAAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 931
OY      216 PheAlaAspLeuAspIleLeuAlaThrGlySerValValGlyGluValSerHisAspPhe 235
DB      932 CTGACGAGCTGAGATCTTTGCTATA-----ATCTTCCTCAGCTGCCATCATGACTGAC 985
OY      236 AsparGlyTrrPheLysSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsn 255
DB      986 GAGCATACCGGAGACACACACAAATTCACATTCAGACTCGG-----1027
OY      256 IleGlyGlyGlyLeuGluAlaLeuGlyTyrAsnAspGluThr-----SerArgHis 272
DB      1028 -----TGTGATCCAGACTATTTCTGTATATGACAGATCTGTACTGGAGATACACCAT 1078
OY      273 AlaLeuLeuArgTyrArgGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThr 292

```

```

DB      1079 TTAAGTCAGACTTATGCGCTTCTGCAAGATGACGAGAAATGATATTTTGAATTAACCTC 1138
OY      293 GlyArgIleAspTrrPheGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGly 312
DB      1139 TCAAAGATGACTGGAGGAGAGTTCGAACTTGATAT-----GAAATGCTGATAGGCC 1192
OY      313 LeuAspArgAspArgArgLysProProIleAlaGlyValGluGlnAspAlaLeuGln 332
DB      1193 ACAGATATGCTTGTGCTACTTCCAAATCAAGCA-----ATGAAGACTGCTGTGACGACAG 1249
OY      333 ProGluLys-----335
DB      1250 CCAGAAAGCATGAAAGCCAAAGCCATTATCCCTATGCTGCAATACAGACATATTAGC 1309
OY      336 -----SerValTyrLeuValSerProTyr 343
DB      1310 CATCCAGCAAAAGCATGGAGCTCATCATCGCTGACAAATGCTCATCTCTGGAGAGATTG 1369
OY      344 PheValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 363
DB      1370 TTC-----AGACAGGCT---GACAGAGAGCAGAGCTG-----GGCTGCTT 1408
OY      364 ValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHisSerGlyTyr 383
DB      1409 TTTTCTCTCTGTGTGACCGA---AAGTCCACTATGCTTGCTGACAGTACAGATGATTTC 1465
OY      384 ValLysTyr---ArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnPro 402
DB      1466 ATTGATTTCATCGTGGAAACCCCTTCACTGCTTACGAGATGACGCAAGAAATTTG 1525
OY      403 AsnHisAlaValProAlaThrLysAspLysGlyLeuThrGly-----SerSerVal 419
DB      1526 AGTCCATTATGATGACAAACCTCTCAACTGGTGGACAGACAGAGCGTTTCGATTTG 1585
OY      420 ThrSerLeuHisAlaLysThrPheIleValAlaSerGlyLysArgIlePheIleGlySerPhe 439
DB      1586 AATAGCATGAGCTGCGCA-----GATGCCAAGCAGTACAGGTCTCAAGACCTCT 1633
OY      440 AsnLeuAspProArgSerAlaArgLeuAsnThrGluMetGlyVal-ValIleGluSerPhe 459
DB      1634 GGTTCAGAGGAAAGTCCCGCATCAAAATTTCTGCTATCTGCTGATCAATGAAGCTTT 1693
OY      459 OlysIleAlaGluGlnMetGluArgThrLeu-----AlaAspThrSerProGluTyrAl 477
DB      1694 AAGCTACTGAGCAGAGAGAGTGTGCAATTCGAGAGAGTGGAGGCGCAAGTACCC 1753
OY      477 ATyrArgValThrLeuAspArg-----HisAsnArgLeuGlnTrrPheHisAspProAlaTh 495
DB      1754 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1813
OY      495 rArgLysThrTyrProAsnGluProGluAlaLysLeuTrrPheLysArgIleAlaLysIle 515
DB      1814 CAAAGGAAATG-----AAGCCAAAGCCAGGCTGGAAGAGGCGCATCTGGCAAGACT 1867
OY      515 e-----LeuSerLeuLeuProIleGluSerLeu 524
DB      1868 GAGAAAGAGCTGTGGAGAAACTTAAGAACTAAGTCAATG 1907

RESULT 36
US-10-010-943-3
; Sequence 3, Application US/10010943
; Patent No. US20020146686A1
; GENERAL INFORMATION:
; APPLICANT: Cook, James William
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; FILE REFERENCE: MNT-197
; CURRENT APPLICATION NUMBER: US/10/010, 943
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: 60/254, 037
; PRIOR FILING DATE: 2000-12-07

```



```

: Sequence 71, Application US/09813153
: Publication No. US20030045459A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: 67 Human secreted proteins
: FILE REFERENCE: P2023
: CURRENT APPLICATION NUMBER: US/09/813,153
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: US/09/363,044
: PRIOR FILING DATE: 1999-07-29
: PRIOR APPLICATION NUMBER: 60/073,160
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,159
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,165
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,164
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,167
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,162
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,161
: PRIOR FILING DATE: 1998-01-30
: PRIOR APPLICATION NUMBER: 60/073,170
: PRIOR FILING DATE: 1998-01-30
: NUMBER OF SEQ ID NOS: 298
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 71
: LENGTH: 1948
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-813-153-71

Alignment Scores:
Pred. No.: 0.0845 Length: 1948
Score: 102.50 Matches: 53
Percent Similarity: 39.42% Conservative: 29
Best Local Similarity: 25.48% Mismatches: 75
Query Match: 3.78% Indels: 51
DB: 9 Gaps: 9

US-10-066-551-4 (1-525) x US-09-813-153-71 (1-1948)
OY 284 SerProLeuTYrGlnLysIleGlnThrGlyAlaSerProGlnSerValGlnThrArg 303
DB 370 AGCCCTTGAGAGCAGAGGCGCAGCAGAGGAGCTCTGCGACCTTCTCTTGGA 429
OY 304 LeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysProIleAla 323
DB 430 AGCATCCGCCAGACCTGCATCTGCAGCGGCGACCCCTGCGCCAGCTTG6GCCAG 489
OY 324 GlyArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValTYrLeuValSerProTYr 343
DB 490 GCCTGCGTCAG--CTGCTGACACTGCCAGAGAGGCTCCAGCTTCACACTAC 546
OY 344 PheValProThr-----LysSerGlyThrAsp 352
DB 547 TGGTCCCTCACAGGCGCTGACATCGGGGCAACGACTGCTCTCCAGCTGGAGAGGCT 606
OY 353 AlaLeuAlaLysLeuValGln-----AspGlyIleAspValThrValLeuThrAsn 369
DB 607 CTTCGCAAGAGCTCAGAGCTGCTGGGCAAGACATTTCTGCTGGCTGCGCACAC 666
OY 370 Ser-----LeuGlnAlaThrAsp-----ValAlaAlaValHisSerGlyTYr 383
DB 667 AGCCCGACACTGGCCAGGACATCCACGCTGACAGTTCTGCGCCGAGTGCCCAT 726
OY 384 ValIleTYrArgLysProLeu-----LeuLysAlaGlyIleLysLeuTYrGlnLeuGln 401
DB 727 GTACGA--CAAGTCCCGCTGGGGCGGCTCACCATGGGTTGTT----- 765
OY 402 ProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSer 421

```

```

DB 765 ----- 765
OY 422 LeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeu 441
DB 766 TTGCACATCCAAATTTCTGTGTGATGAGCGGCAACATATACATGCGCAGCAACATG 825
OY 442 AspProArgSerAlaArgLeuAsnThrGluMetGlyValIle-----GlySerProLys 460
DB 826 GACTGGCGGTCTCTGACGAGAGGTGAAGAGCTTGGCGCTGCTCATATACATGCAAC 885
OY 461 IleAlaGluGlnMetGluArgTYr 468
DB 886 CTGGGCGCAAGACCTGAGAGAGACC 909

RESULT 39
US-10-245-103-41
: Sequence 41, Application US/10245103
: Publication No. US20030068778A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Eaton, Dan
: APPLICANT: Flivaioff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gurney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Phillippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630R1C112
: CURRENT APPLICATION NUMBER: US/10/245,103
: CURRENT FILING DATE: 2002-09-17
: PRIOR APPLICATION NUMBER: 10/197942
: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/08478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: PRIOR FILING DATE: 1998-06-25
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 41
: LENGTH: 1964
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-245-103-41

Alignment Scores:
Pred. No.: 0.0856 Length: 1964
Score: 102.50 Matches: 53
Percent Similarity: 40.57% Conservative: 33
Best Local Similarity: 25.00% Mismatches: 71
Query Match: 3.78% Indels: 55
DB: 9 Gaps: 11

```

```

US-10-066-551-4 (1-525) x US-10-245-103-41 (1-1964)
QY 284 SerProLeuTYrGlnLysIleGlnThr---GlyArgIleAspTrpGlnSerValGlnThr 302
    ||||| : : : : : |||
Db 370 AGCCAGCTTGGAGCCCTGGAGCAGAGCCAGCAGCAGAGGAGCTCTGCCAGCT 429
QY 303 ArgLeuIleSerAspAspProAlaLysGlyLeuAspArgArgLysPro----- 320
    ||||| : : : : : |||
Db 430 GTCTTGTGGAAAGCATCCC---CAGAGCTGCGCATCTGCAGCCGCGAGCCCTCTGCC 486
QY 321 ---ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValIleu 339
    ||||| : : : : : |||
Db 487 CAGCTCTGGGCGCAGCCTGGCTGCAGCTGCAGCTGCAGCTGCAGAGAGCTCCACGTG 546
QY 340 ValSerProTYrPheValProThr-----Lys 348
    ||| : : : : : |||
Db 547 GCTTCATCTACTGCTGCTGCTGCAGAGCTGCAGCTGCAGAGCTGCTCTCCAG 606
QY 349 SerGlyThrAspAlaLeuAlaLysIleValGln-----AspGlyIleAspValThr 365
    ||||| : : : : : |||
Db 607 CTGGAGAGGCTCTTCTGCAGAGCTGCAGCTGCTGGCGAGAGACATTTCTCTGCT 666
QY 366 ValLeuThrAsnSer-----LeuGlnAlaThrAsp-----ValAlaAlaVal 379
    ||| : : : : : |||
Db 667 GTGGCCACACAGCAGCCGCACTGGCCAGACATCCACCGAGCTGCAGTTCCTGCTGCC 726
QY 380 HisSerGlyTYrValLysTYrArgLysProLeu-----LeuLysAlaGlyIleLysIleu 397
    : : : : : : : : : : : |||
Db 727 CGAGGTGCCCATGTACGA---CAGGTGCCCATGTGGCGGCTGCACCGAGGCTGTT 777
QY 398 TYrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySer 417
Db 777 ----- 777
QY 418 SerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGly 437
    ||||| : : : : : |||
Db 778 -----TTCACATCCAAATTCTGGGTTGTGATGAGCGACATATACATGAGCC 825
QY 438 SerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMetGlyValIle--- 456
    ||| : : : : : |||
Db 826 AGTGCACACATGTGAGCTGGGCTCTGCAGCGCAGTGAAGAGCTTGGCGCTCATCTAT 885
QY 457 GluSerProLysIleAlaGlnIleMetGluArgThr 468
Db 886 AACTGACAGCCACTGGCCCAAGACTGGAGAGAGCC 921

RESULT 40
US-10-245-107-41
: Sequence 41, Application US/10245107
: Publication No. US2003006879A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin
: APPLICANT: Eaton, Dan
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Goddard, Audrey
: APPLICANT: Grimaldi, J. Christopher
: APPLICANT: Gutney, Austin
: APPLICANT: Smith, Victoria
: APPLICANT: Stephan, Jean-Phillippe
: APPLICANT: Watanabe, Colin
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: APPLICANT: Fong, Sherman
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: FILE REFERENCE: P3630R1C71
: CURRENT APPLICATION NUMBER: US/10/245,107
: CURRENT FILING DATE: 2002-09-16
: PRIOR APPLICATION NUMBER: 2002-07-18
: PRIOR FILING DATE: 2002-07-18
: PRIOR APPLICATION NUMBER: 60/059114
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/063046

```

```

: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/065027
: PRIOR FILING DATE: 1997-11-10
: PRIOR APPLICATION NUMBER: 60/079689
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/086478
: PRIOR FILING DATE: 1998-05-22
: PRIOR APPLICATION NUMBER: 60/087607
: PRIOR FILING DATE: 1998-06-02
: PRIOR APPLICATION NUMBER: 60/089801
: PRIOR FILING DATE: 1998-06-18
: PRIOR APPLICATION NUMBER: 60/090557
: PRIOR FILING DATE: 1998-06-24
: PRIOR APPLICATION NUMBER: 60/090689
: PRIOR FILING DATE: 1998-06-25
: Remaining Prior Application data removed - See file wrapper or PALM.
: NUMBER OF SEQ ID NOS: 116
: SEQ ID NO 41
: LENGTH: 1964
: TYPE: DNA
: ORGANISM: Homo Sapien
US-10-245-107-41

Alignment Scores:
Score: No. 0.0856 Length: 1964
Percent Similarity: 102.50 Matches: 53
Best Local Similarity: 40.57% Conservative: 33
Query Match: 25.00% Mismatches: 71
DB: 3.78% Indels: 55
Gaps: 11

US-10-066-551-4 (1-525) x US-10-245-107-41 (1-1964)
QY 284 SerProLeuTYrGlnLysIleGlnThr---GlyArgIleAspTrpGlnSerValGlnThr 302
    ||||| : : : : : |||
Db 370 AGCCAGCTTGGAGCCCTGGAGCAGAGCCAGCAGAGGAGCTCTGCCAGCT 429
QY 303 ArgLeuIleSerAspAspProAlaLysGlyLeuAspArgArgLysPro----- 320
    ||||| : : : : : |||
Db 430 GTCTTGTGGAAAGCATCCC---CAGAGCTGCGCATCTGCAGCCGCGAGCCCTCTGCC 486
QY 321 ---ProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlyLysSerValIleu 339
    ||||| : : : : : |||
Db 487 CAGCTCTGGGCGCAGCCTGGCTGCAGCTGCAGCTGCAGCTGCAGAGAGCTCCACGTG 546
QY 340 ValSerProTYrPheValProThr-----Lys 348
    ||| : : : : : |||
Db 547 GCTTCATCTACTGCTGCTGCTGCAGAGCTGCAGCTGCAGAGCTGCTCTCCAG 606
QY 349 SerGlyThrAspAlaLeuAlaLysIleValGln-----AspGlyIleAspValThr 365
    ||||| : : : : : |||
Db 607 CTGGAGAGGCTCTTCTGCAGAGCTGCAGCTGCTGGCGAGAGACATTTCTCTGCT 666
QY 366 ValLeuThrAsnSer-----LeuGlnAlaThrAsp-----ValAlaAlaVal 379
    ||| : : : : : |||
Db 667 GTGGCCACACAGCAGCCGCACTGGCCAGACATCCACCGAGCTGCAGTTCCTGCTGCC 726
QY 380 HisSerGlyTYrValLysTYrArgLysProLeu-----LeuLysAlaGlyIleLysIleu 397
    : : : : : : : : : : : |||
Db 727 CGAGGTGCCCATGTACGA---CAGGTGCCCATGTGGCGGCTGCACCGAGGCTGTT 777
QY 398 TYrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySer 417
Db 777 ----- 777
QY 418 SerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGly 437
    ||||| : : : : : |||
Db 778 -----TTCACATCCAAATTCTGGGTTGTGATGAGCGACATATACATGAGCC 825
QY 438 SerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMetGlyValIle--- 456
    ||| : : : : : |||
Db 826 AGTGCACACATGTGAGCTGGGCTCTGCAGCGCAGTGAAGAGAGCTTGGCGCTCATCTAT 885

```



Tue May 20 09:34:01 2003

us-10-066-551-4.rnpb

Page 39

```

Qy      457  GLuserProlysileAglugImetGluayThr 468
          :::::  :::::
Db      886  AACTGCAGCCACCTGGCCCAAGACCTGGAGAAGACC 921

```

Search completed: May 19, 2003, 19:13:32  
Job time : 559 secs



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 19, 2003, 16:51:32 ; Search time 1426 Seconds

(without alignments)  
5962.574 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 2713

Sequence: 1 MRANKTQAMPSETISLTKT.....KLMKRIAKIISLPISL 525

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+pgn.model -DEV=xlh

-Q/cgpn2.1/USPFO\_SPOOL/US1006551/runat\_12052003\_091143\_22366/app\_query.fasta\_1.711

-DB=EST -QFMT=fastlap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0

-UNIT3-bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi -LIST=45

-DOCALLGN=200 -THR\_SCORE=200 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=40 -MODE=LOCAL

-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US1006551 @CGN 1.1 1456 @runat\_12052003\_091143\_22366 -NCPU=6 -ICPU=3

-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120

-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estinu:\*

4: em\_estinu:\*

5: em\_estlo:\*

6: em\_estlo:\*

7: em\_estro:\*

8: em\_estro:\*

9: gb\_estl:\*

10: gb\_estl:\*

11: gb\_estl:\*

12: gb\_estl:\*

13: gb\_estl:\*

14: gb\_estl:\*

15: em\_estfun:\*

16: em\_estfun:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_hum:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_vrt:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	326	12.0	706	13	BJ030047	BJ030047 BJ030047
2	303	11.2	571	10	BE105595	BE105595 UT-R-BXO-
3	294	10.8	470	9	AI385165	AI385165 UT-R-E0-C
4	238	8.8	373	9	AA546779	AA546779 VK02C08.F
5	181.5	6.7	315	9	AI363691	AI363691 AGWMEFCAR
6	161.5	6.0	614	17	BH382846	BH382846 AG-ND-125
7	154	5.7	661	17	AO156942	AO156942 nbx00008N
8	145.5	5.4	366	12	BF544011	BF544011 UT-R-E0-C
9	125	4.6	95	9	AA669619	AA669619 ac19f05.s
10	121	4.5	713	13	BI269552	BI269552 NF005A10I
11	120	4.4	682	14	BO153318	BO153318 NF033H12I
12	118.5	4.4	549	13	BM088538	BM088538 502040 MA
13	115.5	4.3	543	10	AM655356	AM655356 105979 MA
14	115	4.2	704	9	AU212487	AU212487 AU212487
15	114	4.2	570	14	D79520	D79520 HM272608B
16	113	4.2	1109	10	BE283121	BE283121 601103580
17	112.5	4.1	407	17	TA131A08Q	TA131A08Q T. brucei
18	111	4.1	676	13	BI103788	BI103788 602887931
19	110	4.1	548	9	AA207840	AA207840 mw81B08.F
20	108.5	4.0	765	13	BC920511	BC920511 602824841
21	108	4.0	496	12	BF638093	BF638093 NF043F08P
22	108	4.0	545	10	AM657282	AM657282 110034 MA
23	108	4.0	629	9	AU211080	AU211080 AU211080
24	108	4.0	637	13	BJ144950	BJ144950 BJ144950
25	108	4.0	649	13	BJ130081	BJ130081 BJ130081
26	108	4.0	696	13	BM600244	BM600244 170006870
27	108	4.0	955	12	BF206365	BF206365 601869921
28	107.5	4.0	440	10	AW437576	AW437576 78841 MAR
29	105.5	3.9	912	13	BI522691	BI522691 603175730
30	105	3.9	632	10	BE332913	BE332913 us96a10.Y
31	104.5	3.9	366	10	AM030783	AM030783 EST274038
32	104.5	3.8	628	9	AU218786	AU218786 AU218786
33	104	3.8	730	13	BJ152678	BJ152678 BJ152678
34	103.5	3.8	681	17	AQ647885	AQ647885 RPC193-Ec
35	103.5	3.8	1333	13	BI856614	BI856614 603385754
36	103	3.8	857	13	BI772082	BI772082 603059160
37	102.5	3.8	533	13	BJ326903	BJ326903 BJ326903
38	102.5	3.8	573	13	BJ328796	BJ328796 BJ328796
39	102.5	3.8	607	14	BO387515	BO387515 NISC_mn24
40	102.5	3.8	685	13	BJ282990	BJ282990 BJ282990
41	102	3.8	847	9	AA391000	AA391000 LD09965.5
42	102	3.8	1071	14	BM919101	BM919101 AGENCOURT
43	102	3.8	1197	17	AF029619	AF029619 AF029619
44	101	3.7	295	10	AM435575	AM435575 74086 MAR
45	101	3.7	665	13	BJ154883	BJ154883 BJ154883

## ALIGNMENTS

RESULT 1  
BJ030047/c 706 bp mRNA linear EST 05-DEC-2001  
LOCUS BJ030047 NIBB Mochli normalized Xenopus neurula library Xenopus  
DEFINITION laevis cDNA clone XL003a02 5', mRNA sequence.  
ACCESSION BJ030047  
VERSION BJ030047.1 GI:17371471  
KEYWORDS EST.  
SOURCE African clawed frog.  
ORGANISM Xenopus laevis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
Xenopodinae; Xenopus.  
REFERENCE 1 (bases 1 to 706)  
Kitayama,A., Terasaka,C., Mochli,M., Ueno,N., Shii,T. and Kohara

TITLE  
JOURNAL  
COMMENT

Expressed genes in X. laevis embryo  
Unpublished (2001)  
Contact: Tadasu Shin-I  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tschini@genes.nig.ac.jp

FEATURES  
source

1. 706  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="X1003a02"  
/clone\_lib="NIBB Mochii normalized Xenopus neurula library"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 15"  
/note="Vector: pBSRN3; Site\_1: NotI; Site\_2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library is subtracted and was constructed by N. Garrett and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 198 a 172 c 154 g 181 t 1 others

ORIGIN

Alignment Scores:  
Pred. NO.: 9.4e-28 Length: 706  
Score: 326.00 Matches: 71  
Percent Similarity: 56.00% Conservative: 41  
Best Local Similarity: 35.50% Mismatches: 74  
Query Match: 12.02% Indels: 14  
DB: 13 Gaps: 3

US-10-066-551-4 (1-525) x BU030047 (1-706)

OY 199 GlyGlyArgAsnIleGlyAspGluThrPheValGlyGluAspThrValPheAlaAsp 218  
DB 705 GGAGAGACAAATATGGTATGCTATTTGGAGCAGGGAGAGCCACTTTTTCGGAT 646  
OY 219 LeuAspIleLeuAlaThrGlySerValAlaGlyGluValSerHisAspPheAspArgTyr 238  
DB 645 TTAGATGTCATGGCAATAGACCCGCGTAGAGACGTTGCCGATTTCCCGCCTAC 586  
OY 239 TrpAlaSerHisSerAlaHisAsnAlaThrArgIleIleArgSerGlyAsnIleGlyLys 258  
DB 585 TGGTATTCGCAATCGCTTACCCCTTACAGACGTCCTG-----GATGTCCCGGAG 535  
OY 259 Gly-----LeuGlnAlaLeuGlyTyrAsnAspGluThrSerArg 271  
DB 534 GGTGAATGGCGGATCGCATGCTACCCGCTCTCGCATACGATCCATGACGAT 475  
OY 272 HisAlaLeuLeuArgTyrArgIleThrValGlnGlnSerProLeuTyrGlnLysIleGln 291  
DB 474 -----CGTTATTACGCAAAATGCAATCCAGTCCATTATTAATCATCTGGTT 427  
OY 292 ThrGlyArgIleAspTyrPheIleValGlnThrArgLeuIleSerAspProAlaLys 311  
DB 426 GATGGAACATTCGCGCTTATCGGCGAAGACACGTTTATTAAGATGATCCGGGANA 367  
OY 312 GlyLeuAspArgAspArgTyrGlySerProIleAlaGlyArgLeuGlnAspAlaLeuLys 331  
DB 366 GGGGAGGCGCAAGCAACGCGCATTCATCGCAGCGCGCTGTCATATCATGGCC 307  
OY 332 GlnProGlyLysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThr 351  
DB 306 TCACCCAGTGAACGATGATATTAATCTCTCTATTGTTTGAACGACAGCCAGGTGTC 247  
OY 352 AspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeu 371  
DB 246 GCGCAACCTTATCGGATGTGTGAGAAAGGGGTAAAGATTGCGATCTCAACCAATTCTCT 187

OY 372 GlnAlaThrAspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeu 391  
DB 186 GCCGTAATATACGAATATGCTTCATACGATTTCTTATTTATTAATCATCTGTTCT 127

RESULT 2

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

BE105595 571 bp mRNA linear EST 13-JUN-2000  
UI-R-BX0-ash-g-02-0-UI.s1 UI-R-BX0 Rattus norvegicus cDNA clone  
UI-R-BX0-ash-g-02-0-UI 3', mRNA sequence.  
BE105595  
BE105595.1 GI:8497697  
EST.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 571)  
Bonaldi,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
97044477  
Contact: Soares, MB  
Program for Rat Gene Discovery and Mapping  
451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: msoares@iue.uiowa.edu  
Oligo-dT track not found, Not I site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: clones will be available through  
Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA-No.

FEATURES  
source

1. 571 Location/Qualifiers  
/organism="Rattus norvegicus"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-BX0-ash-g-02-0-UI"  
/clone\_lib="UI-R-BX0"  
/dev\_stage="embryonic 13 dpc"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; The UI-R-BX0  
library is derived from 13 dpc whole embryo tissue. For a  
detailed description of the library from which this clone  
was derived, please visit our web site at  
rategen.eng.uiowa.edu.  
TAG\_SEQ=None found"

BASE COUNT 81 a 213 c 193 g 84 t

ORIGIN

Alignment Scores:  
Pred. No.: 3.52e-25 Length: 571  
Score: 303.00 Matches: 68  
Percent Similarity: 58.72% Conservative: 33  
Best Local Similarity: 39.53% Mismatches: 64  
Query Match: 11.17% Indels: 8  
DB: 10 Gaps: 2

US-10-066-551-4 (1-525) x BE105595 (1-571)

OY 284 SerProLeuTyrGlnLysIleGlnThrArgIleAspTyrPheIleValGlnThrArg 303  
DB 65 TCCCGCTGGAGAGAGCTTCGCGCGCGCTGACGCTGACTGGGCCCGCTGCAGC 124  
OY 304 LeuIleSerAspProAlaLysGlyLeuAspArgAspArgLysProProIleAl 323  
DB 125 CTGCTGTCGATGACCTTC-GAA-----GATCCGACCGAAGCGACCCGCGCTGCAC 177

QY 323 aGlyArgLeuGlnAsp-----AlaLeuYsgInProGlyLysSerValTyrIle 339  
 Db 178 CGAGACGATCGCCGACGATGTGGCGGACCTGCGGACAGACGACGAGCCGTCAT 237  
 QY 339 uValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAlaValAla 359  
 Db 238 CATCTCGCCCTACTACGTGCGCGCGCGCGCATGCGACTCTGACAGAGACTCGCGGC 297  
 QY 359 nAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAla 379  
 Db 298 CAGGCGCGCGCGCGGTGAGGTGTGACCACTGCGCGCGCGCGCGCGCGCGGTGT 357  
 QY 379 LHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrG 399  
 Db 358 ACACATCGGCTACG 417  
 QY 399 uLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerVa 419  
 Db 418 GATGGAGCGCGCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477  
 QY 419 LThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPh 439  
 Db 478 CGCCACTGTCAGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537  
 QY 439 eAsnLeuAspProArgSerAlaArgLeuAsnThr 450  
 Db 538 GAACTGACCG 571  
 RESULT 3  
 A1385165/c 480 bp mRNA linear EST 27-JAN-1999  
 LOCUS A1385165 UI-R-E0-cg-02-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone  
 DEFINITION UI-R-E0-cg-02-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone  
 ACCESSION A1385165  
 VERSION A1385165.1 GI:4197947  
 KEYWORDS EST.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 480)  
 REFERENCE 1 Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 AUTHORS Normalization and subtraction: two approaches to facilitate gene  
 COMMENT discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 CONTACT: Soares, MB  
 PROGRAM for Rat Gene Discovery and Mapping  
 UNIVERSITY of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 TEL: 319 335 8250  
 FAX: 319 335 9565  
 EMAIL: msoares@iubm.weeg.uiowa.edu  
 Oligo-dT track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.  
 Fatima Bonaldo, Ph.D. Clone distribution: clones will be available  
 through Research Genetics  
 Seq primer: M13 Forward.  
 FEATURES  
 source  
 1..480  
 Location/Qualifiers  
 ..organism="Rattus norvegicus"  
 ..strain="Sprague-Dawley"  
 ..db\_xref="taxon:10116"  
 ..clone="UI-R-E0-cg-02-0-UI"  
 ..clone\_1fb="UI-R-E0"  
 ..dev\_stage="embryonic"  
 ..lab\_host="DH10B (Life Technologies)"  
 ..note="Vector: pUT73D-pac (Pharmacia) with a modified  
 polylinker, Site\_1: NotI, Site\_2: EcoRI; This library  
 consists of a mixture of individually tagged normalized  
 libraries constructed from 8, 12 and 18-day embryo. The  
 tag is a string of 3-5 nucleotides present between the

Not 1 site and the oligo-dT track which allows  
 identification of the library of origin of a clone within  
 the mixture."

BASE COUNT 88 a 176 c 154 g 62 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3,07e-24 Length: 480  
 Score: 294.00 Matches: 65  
 Percent Similarity: 64.34% Conservative: 18  
 Best Local Similarity: 50.39% Mismatches: 46  
 Query Match: 10.84% Indels: 0  
 DB: 9 Gaps: 0

US-10-066-551-4 (1-525) x A1385165 (1-480)

QY 72 GlyLeuSerAspIleTyrLeuLeuAsnAspProHisGluAlaPheAlaAlaAla 91  
 Db 389 GGCCG 330  
 QY 92 LeuIleGluSerAlaGluHisSerLeuAspLeuGlnTyrTyrIleThrPargAsnAspIle 111  
 Db 329 CTGGTCGGCGCGGCGACAGCCACATCATGATGCGCATATATATCTGCGCAACATGTG 270  
 QY 112 SerGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgValArg 131  
 Db 269 GCGTGGCGGATGCTGTTACAGAGCTGGCGGCGCGCGCGCGCGCGCGCGCG 210  
 QY 132 LeuLeuLeuAspAspAsnAsnThrArgGlyLeuAspAspLeuLeuAlaLeuAspSer 151  
 Db 209 CTGCTCTCTGATGATTTCCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCG 150  
 QY 152 HisProAsnIleGluValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeu 171  
 Db 149 CTGCCAGCGCTGACGTGCGGCTGTCAACCGCGCGCGCGCGCGCGCGCGCG 90  
 QY 172 GlyTyrLeuThrAspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAla 191  
 Db 89 AACATGATGTTGATTTTCCCGCGCTCAACCGCGCGCGCGCGCGCGCGCGCG 30  
 QY 192 AspAsnArgAlaThrIleLeuGlyGly 200  
 Db 29 GAGCGCATCGCACGCGTGTACGCGGT 3

RESULT 4  
 AA546779/c 373 bp mRNA linear EST 05-AUG-1997  
 LOCUS AA546779 vk02c08.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA  
 DEFINITION clone IMAGE:945422 5', mRNA sequence.  
 ACCESSION AA546779  
 VERSION AA546779.1 GI:2308070  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 373)  
 REFERENCE 1 (bases 1 to 373)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LML; contact the  
 IMAGE Consortium (info@image.lml.gov) for further information.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AI363691	AI363691.1	GI:115312	EST.	Onchocerca volvulus.
				Onchocerca volvulus
				Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea
				Onchocercidae; Onchocerca.

RESULT	6	
BH382846		
LOCUS		
DEFINITION	BH382846	614 bp DNA linear GSS 10-DEC-2001
	AG·ND-125J3..TTF ND-TAM Anopheles gambiae genomic clone AG-ND-125J3,	
DNA sequence.		
ACCESSION	BH382846	
VERSION	BH382846.1	GI:17328988

KEYWORDS	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
GSS.	African malaria mosquito.				
SOURCE	ORGANISM	Amopheles gambiae	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.		
	1 (bases 1 to 614)	Shetty,J., Mulek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J	Direct Submission of BAC-end sequences from Anopheles gambiae	Unpublished (2001)	
	Other-GSSs: AG-ND-125J3.TR	Contact: Brendan J Loftus	Department of Eukaryotic Genomics	The Institute for Genomic Research	9712 Medical Center Dr., Rockville, MD 20850, USA
	Tel: 301 838 0208				
	Fax: 301 838 3543				
	Email: b.loftus@etlgr.org				
	This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.				
	Seq primer: M13 For				
	Class: BAC ends.				
FEATURES	location/Qualifiers				
source	1..614				
	/organism="Anopheles gambiae"				
	/strain="PEST"				
	/db_xref="taxon:7165"				
	/clone="AG-ND-125J3"				
	/clone_lib="ND-TAM"				
	/note="vector: pECBAC1, Site_1: HindIII"				
BASE COUNT	131 a 203 c 184 g 96 t				
ORIGIN					
Alignment Scores:					
Pred. NO.:	2.37e-08	Length:	614		
Score:	161.50	Matches:	39		
Percent Similarity:	54.63%	Conservative:	20		
Best Local Similarity:	36.11%	Mismatches:	44		
Query Match:	5.95%	Indels:	5		
	17	Gaps:	4		
US-10-066-551-4 (1-525) x BH382846 (1-614)					
QY	421 SerLeuHISAlaLysThrPheIleValAspGlyValArgIlePheIleLysSerPheAsn	440			
	.....	.....			
	1 AGCTGCACAGCAAGCGATGATCTTTGATGCCGAGAACTGTTATATCGCGCTTCAAC	60			
QY	441 LeuAspProArgSerAlaArgLeuAsnThrGluMetGlyValValIleGluSerProLys	460			
	.....	.....			
	61 TTTCAGCCGCGCGTCCGTCTGTGGAACACCGAAGTCGGCGTCTGTAGACAGCCCGAG	120			
QY	461 IleAlaGluIleMetGluArg--ThrIleuAlaAspThrSerProGluTyrAlaTyrArg	479			
	.....	.....			
	121 CTGGCGGAGCATGTGGCAACCTGGCGGTGGAAGCATGGCCCACTTGACATATGAG	180			
QY	480 ValThrLeuAspArgHisAsnArgLeuGlnTrp-----HisAspProAlaThrArgLys	497			
	.....	.....			
	181 GCGAAATTC---CAGGACGGCCCAAGTGTGTGTGGTGGACCGAGATACGCGCACTCCAC	237			
QY	498 ThrTyrProAsnGluProGluAlaLysLeuTyrPlyValGluIleAlaLysIleLeuSer	517			
	.....	.....			
	238 ACCCTGGGCGCGGAGCCG---CGCAGTTGTGTGGCGCGGTTCATACCGGTTCGCCAC	294			
QY	518 LeuLeuProIleGluSerLeuLeu 525				
	.....				
	295 AGCTTGCGCTGTGGAACGATCTCTC 318				

```

RESULT 7
A0156942 661 bp DNA linear GSS 12-SEP-1998
LOCUS A0156942
DEFINITION nxbx0008N19r CUG1 Rice BAC Library Oryza sativa genomic clone
ACCESSION A0156942
VERSION A0156942.1 GI:3553967
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euhartoideae; Oryzeae; Oryza.
1 (bases 1 to 661)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
JOURNAL
COMMENT Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 253.

FEATURES
Source
location/Qualifiers
1..661
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nippobare"
/db_xref="taxon:4530"
/clone="nxbx0008N19r"
/clone_id="CUG1 Rice BAC Library"
/lssue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBel0BAC11; Site.1: HindIII; Site.2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nippobare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT 146 a 156 c 185 g 174 t
ORIGIN

Alignment Scores:
Pred. No.: 2.08e-07 Length: 661
Score: 154.00 Matches: 44
Percent Similarity: 46.32% Conservative: 19
Best local Similarity: 32.35% Mismatches: 61
Query Match: 5.68% Indels: 12
DB: 17 Gaps: 4

US-10-066-551-4 (1-525) x A0156942 (1-661)
Oy 256 IlleglyLysgIleuGlnAlaLeuGcIyTrtAsnAsprGluTrtSerArgHsAlaLeuLeu 275
db 5 ATTGGCAATGGCAGCGCGCGCGCGCGGATATATGCACCATCTGCAGCGGCTGGCG-----58

```

```

OY 276 ArgTyrArgLutThrValGlnGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIle 295
Db 59 -----GGGAGAGGAGGGCGCTGTATCAC-----GGACCGGTCGGCTG 94
OY 296 AsPTrpGlnSerValGlnThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArg 315
Db 95 CGCTGGACCGCA---GAGCTCGCGCATGTGGCCGCGGCGGAAAGCGC---CATATG 148
OY 316 AspArgArgLysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGlnLys 335
Db 149 GCGCGAGCGGACAGACTGCTATGAGACTGCTGATGCCCGCGATGATGATGCCAGCGG 208
OY 336 SerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeuAla 355
Db 209 CATCTGCCCATCATCTCCCTATTTCATTCCTCCGCGGAGAGCGGATGAAGCGCTTGGC 268
OY 356 LysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaThrAsp 375
Db 269 GCGCTACGACGACGCGGGGCGGAGTGGCGGTGACCAATTGCTGCGCGTACAAAC 328
OY 376 ValAlaValAlaHisSerGlyTyrValLysTyrArgLysProLeuLeu 391
Db 329 GTCCCGCGCGTGTAGAGGACATTACATCATCTGCTAGGCTCTT 376

RESULT 8
BF544011 366 bp mRNA linear EST 11-DEC-2000
LOCUS UI-R-E0-cg-9-02-0-UI.r1 UI-R-E0 Rattus norvegicus cDNA clone
DEFINITION UI-R-E0-cg-9-02-0-UI 5', mRNA sequence.
ACCESSION BF544011
VERSION BF544011.1 GI:11635118
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 366)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@iue.wiwee.uiowa.edu
cDNA Library Preparation: M.B. Soares Lab Clone Distribution:
clones will be available through Research Genetics (www.resgen.com)
This clone is also available through the I.M.A.G.E. Consortium at
LMNL (info@image.llnl.gov). IMAGE ID= 1791952
Seq primer: M13 Forward.
Location/Qualifiers
1..366
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-E0-cg-9-02-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: NotI; Site_2: EcoRI. This library
consists of a mixture of individually tagged normalized
libraries constructed from 8, 12 and 18-day embryo. The
tag is a string of 3-5 nucleotides present between the
Not I site and the oligo-dT track which allows
identification of the library of origin of a clone within
the mixture."

```

```

BASE COUNT 39 a 125 c 138 g 64 t
ORIGIN
Alignment Scores:
Pred. No.: 7,87e-07 Length: 366
Score: 145.50 Matches: 45
Percent Similarity: 53.60% Conservative: 22
Best Local Similarity: 36.00% Mismatches: 42
Query Match: 5.36% Indels: 16
DB: 12 Gaps: 5

US-10-066-551-4 (1-525) x BF544011 (1-366)
OY 19 LysThrArgSerLeuIleSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 38
Db 14 CGATGCGCGCGGCTGATCGCGCTTGGCCATCATCATCGCTCC-----ACGTGG--- 61
OY 39 ProProLeuGluGluArgThrGlnSerArgHisPheAsnThrSerLysProValArgLeu 58
Db 62 ---CCGATGCGCGCGCGGACG---CCTGCGCGTGTGCGCGGCGCATCGATCATCTGC 115
OY 59 AspAsnIleLeuGlnIleArgHis-Thr-----ProHisTh 70
Db 116 CCATCTTTGTGCGGCTGCGCGGCGATGACGCGTGGCCAGCTGTGCGATGCGCATCC 175
OY 70 RasnGlyLeuSerAspIleTyrLeuLeuAsnAspProHisGlnAlaPheAlaAlaArgAl 90
Db 176 G---GGCGCCAGCGCGCGGCGAGTGTGATGATGAGCGCGCGCGCGTGGCGCAT 232
OY 90 AlaLeuIleLeuSerAlaGlnHisSerLeuAspLeuIleTyrTyrIleTyrPargAsnAs 110
Db 233 CTGCTGTGTGCGCGCGGCGGACGACGCGCATGATGATGATGATGATGATGATGATGATG 292
OY 110 PileserGlyArgLeuLeuPheAsnLeuValTyrLeuAlaAlaGluArgGlyValArgVa 130
Db 293 TGTGGCGGCGCGCGCTGCTGTTAGACAGCTGCGGACCGGCGCGGCGCGGCGTCCGT 352
OY 130 LArgLeuLeuLeu 134
Db 353 GCGGCTGCTGCTC 365

RESULT 9
AA669619 95 bp mRNA linear EST 20-NOV-1997
LOCUS ac19f05.s1 Stralagene ovary (#937217) Homo sapiens cDNA clone
DEFINITION IMAGE:856929 3', mRNA sequence.
ACCESSION AA669619
VERSION AA669619.1 GI:2631118
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 95)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M., Martin
J., Moore,B., Schellenberg,R., Steptoe,M., Tan,F., Theisling,B.,
White,Y., Wyllie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
JOURNAL MEDLINE
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LMNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham.
Location/Qualifiers
1..95
/organism="Homo sapiens"
/db_xref="taxon:9606"

```



BASE COUNT	22 a	20 c	25 g	28 t
ORIGIN				

Alignment Scores:	
Pred. No.:	2.23e-05
Score:	125.00
Percent Similarity:	80.65%
Best Local Similarity:	74.19%
Query Match:	4.61%
DB:	9
Length:	99
Matches:	80
Conservative:	2
Mismatches:	6
Indels:	0
Gaps:	0

US-10-066-551-4 (1-525) x AA669619 (1-95)

Oy 177 PheProArgLeuAsnArgAlaGlyMetHisAsnLysSerPheThrAlaAspAsnArgAlaThr 196  
||| ||||||||||||||||||||| ||| |||  
Db 3 TTTTCCCGCTTAATCGCCGATGCACAATAAAGTTCACTGTGGATGGCGTGTCACC 62

```

Qy 197 ILEuGLyGLyArgAsnILEgLyAspGLuTyR 20
      ::::::::::::::::::::::::::::||
Db 63 CTGCTGGGAGACGAATATTGGTGATGCCCTAT 95

```

RESULT 10	LOCUS	DEFINITION
B1269552	713 bp	linear EST 18-JUL-2001
B1269552	713 bp	linear EST 18-JUL-2001
NC005101.R11072	Irradiated Medicago truncatula cDNA clone	
NC005101.R5	5' mRNA sequence.	

ACCESSION BI269552 GI:14876361  
VERSION BI269552.1  
KEYWORDS EST.

SOURCE	ORGANISM
barrel medic.	<i>Medicago truncatula</i>
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
	Rosidae; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
	<i>Medicago</i> .

REFERENCE 1 (bases 1 to 713)

**AUTHORS**  
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

**TITLE** Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula irradiated library  
**JOURNAL** Unpublished (2001)

COMMENT      Contact: May GD

ology Division  
el Roberts Noble Foundation  
Noble Parkway, Ardmore, OK 73402, USA

Insert Length: 713 Std Error: 0.00  
Plate: 005 row: A column: 10  
Seq primer: TCACACAGGAACACGCTATGAC.

FEATURES	Location/Qualifiers
source	1. .713

```

/organism="Medicago truncatula"
/db_xref="cazon:3880"
/clone_id="NF05A101R"
/clone_lib="Irradiated"
/tissue_type="Seedlings"
/dev_stage="Seedling"
/note="Vector: Lambda Zap; Seedlings were exposed either to 100 Gy gamma or 0.5, 1, 5, or 10 kJ/m2 UV irradiation
gamma-irradiated samples were harvested at 6, 12, 24 and

```

48 hours after treatment. UV-irradiated samples were harvested 24 hours post-treatment. CDNA was prepared from polyA<sup>+</sup> enriched, pooled samples of equivalent amounts of total RNA from each sample. The CDNA was directionally ligated into the uni-zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant uni-zap XR vector using Exassist helper phage and the E. coli strain XL1-Blue MRF<sup>'</sup> (Stratagene). Excised plasmids were plated using SOLR cells."

	a	c	g	t	others
BASE COUNT	164	181	196	157	15
ORIGIN					

Alignment Scores:	
Pred. No.:	0.00193
Score:	121.00
Percent Similarity:	39.59%
Best Local Similarity:	24.87%
Query Match:	4.46%
DB:	13
length:	711
Matches:	429
Conservative:	29
Mismatches:	94
Indels:	25
Gaps:	6

US-10-066-551-4 (1-525) x BI269552 (1-713)

```

QY 93 IIEGLUSERLAIGLHHISSEIREUASPLEUGLINTYRTYRIETPRARGASNAPILESER 112
    ||| |||:: :: |||:: ::||| |||
Db 15 ATTGCGAGGCACACAAGACGATCATCTTGGAAACGTTTATCTGTGTTTGAGAGATGACGTC 74

```

QY 113 GLYARGLEULEU-PHEASNLEUVALTYRLEUALAAGLARGGLVALARGVALARGLE 133  
:::  
DB 75 GGCAACAACACTGCATGCGGGCACCTACTGTGGCAGCAGCCGACACGCCGGGGTTAAAGCCGAATT 133

QY	132	uLeuEnuAspAspAsnAaTTrArgLYLeuAspLeuLeuAlaLeuAspSerH	15
Db	135	CTTGCTGGATGGCTACGGTTGCGCGGATCTACGGCGATAGTTTGTCATAGACATGACGGC	19
QY	152	sProAsnIleGluValArgLeuPheAsnProPheValLeuArgLYsTrpArgAlaLeuGI	17

```

Db      195  AGCTGGCGTAGTGTTCGCTACTACGATCCC-----CGCCCTCGCCTTTTGG 242
          :::      |||      ::::: |||

```

Qy	172	yTyrIleuThraspHePrcOAgLeuAsnArgArgmethIsmsIysSerPhetHraIAs	199
Db	243	TATGGCACCAT-----GTGTTTCGGCGATGCACAAATTTGTCGATCGA	293
Qy	192	pAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspCulTyrPheIysValGlyCl	211

Db 294 CGCGGTATAGCCTTATTGGCGGCTGAATTACTCCGCCGAGCATATGTCACGCTACGG 353

QY 212 uasPThValphealaspLeuAlaThrGlySerValValGlyGluValSe 23

Db 354 TCCAGAGCCTAAACACGAGATTACCGCGTTACGCCCTTGAAAGGCCGATTTGTCSAAGATATTCT 411

QY 232 rHisasprheasr-----ArgTytTpralase 241

Db 414 C---CAGTTGAGCTGGAAACCTGCCCTGACAGAGCGGGCAGCGCGTGGCGACG 470

QY	241	rhissrslahlsnslatlrarglleilear	gsrgclgylasmlleglysglyleu	--	266
			:::		
Db	471	TCAT-----CACCAAGCGGAAGAGAACCCGCGCGGAGAGCGCANTATTGCTGGT	522		

QY 261 -----GlnAlaLeuGlyTyrAsnAspGluThrSerArgHisAlaLeu 274  
::: |||::: ||||| |||

Db 525 CTGGCGGATACNAANACATCGCGATGATATTGAACGCCATTATTG 573

RESULT 11  
BQ153318

LOCUS	BQ15318	682 bp	mRNA	linear	EST 24-APR-2002
DEFINITION	NF033H12IR1F1102 Irradiated Medicago truncatula cDNA clone				

NE033H12IR 5', mRNA sequence

ACCESSION BQ153318

REVISIONS

REV	DESCRIPTION	DATE
01	INITIAL RELEASE	20290377

KEYWORDS	EST.
COLLAPSE	barrel modic

ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;  
Medicago.  
1 (bases 1 to 682)  
Torres-Jerez, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula irradiated library  
Unpublished (2001)  
JOURNAL  
COMMENT Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 682 Std Error: 0.00  
Plate: 033 row: H column: 12  
Seq primer: TCACACAGGAACACGCTATGAC.  
Location/Qualifiers  
1..682  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone="NF033H121R"  
/clone\_lib="Irradiated"  
/tissue\_type="seedlings"  
/dev\_stage="seedling"  
/note="Vector: Lambda zap: Seedlings were exposed either  
to 100 Gy gamma or 0.5, 1, 5, or 10 kJ/m2 UV irradiation.  
Gamma-irradiated samples were harvested at 6, 12, 24 and  
48 hours after treatment. UV-irradiated samples were  
harvested 24 hours post-treatment. cDNA was prepared from  
polyA+ enriched, pooled samples of equivalent amounts of  
total RNA from each sample. The cDNA was directionally  
ligated into the Uni-Zap XR vector (Stratagene) and  
packaged using the Gigapack III Gold packaging extracts.  
Phagemids containing cDNA inserts were in vivo excised  
from the recombinant Uni-Zap XR vector using EXASist  
helper phage and the E. coli strain XL1-Blue MRF'  
(Stratagene). Excised plasmids were plated using SOLR  
cells."

BASE COUNT 154 a 173 c 199 g 152 t 4 others  
ORIGIN

Alignment Scores:  
Pred. No.: 0.00235 Length: 682  
Score: 120.00 Matches: 50  
Percent Similarity: 38.32% Conservative: 32  
Best Local Similarity: 23.36% Mismatches: 101  
Query Match: 4.42% Gaps: 31  
DB: 14 Gaps: 6

US-10-066-551-4 (1-525) x B0153318 (1-682)

OY 93 IlegluserAlagluHlserleuAspleuGlnTyrtTlletPrAgasnasplleser 112  
DB 15 ATTGGCAGGCGCCAGCAAGCGCATCTTCTGAACGTTTATCTGTTGAGAGTACGTC 74  
OY 113 GtATgLeuLeu-PheAsnLeuValTyrtLeuAlaIagluAArgglYvalAArgle 132  
DB 75 GCGCAACAACTGCATGCNGCGACTACTGGCAGCAGCGCAGCGGGTTAAAGCGGAACT 134  
OY 132 uleuLeuAspAsnAsnThraArgglYleuAspAspleuLeuAlaLeuAspSerHl 152  
DB 135 CTTCGCTGATGGCTAGCGTTCGCCGAGTCTCAGCATGTGTTGCAATGAACTGACGCG 194  
OY 132 sTrOAsnIlegluValAArgleuPheAsnPropheValleuAArgglYtTPArAlaleuGl 172  
DB 195 AGCTGCGCTAGTTCGCTACTACGATCC-----CGCCCTCGCCCTTTTTCG 242  
OY 172 yTyrleuThraPheProArgrLeuAsnAArglMethIsAsnLysSerPheThAlaAs 192

DB 243 TATGGCGAACCAAT-----GTCTTCGCCGCGATGATCCCAAAATGTGTGATCGA 293  
OY 192 pAsnArgrAlaThrlleLeuGlYlAArgsnlleglAspglYurPheleYvalGl 212  
DB 294 CCGCGGTATAGCCTTTATTTGGCGGTGAATTATCTCCCGCCAGCAATATCTCAGCTACCG 353  
OY 212 uAsPThraValPheAlaAspleuAspIleuAlaThrglYserValAlglYgluValse 232  
DB 354 TCCAGAGGCTAAACAGANTTACGGGTACCGCTTGAGAGCCGATTTCTGAAATATCT 413  
OY 232 rHisAsPheAsp-----ArgTYrtTPAlase 241  
DB 414 C--CAGTTTGAGCTGGAACACCTCCCTGACAGACGCGCGCAGCGCTGGCGAGC 470  
OY 241 rHisserAlaHisAsnAlaThraArglIleleArgerglYAsnIleglYsglYleuGl 261  
DB 471 TCATCCAAAGCGCGAAGANAACCG-----CAGCGGGAGAGAGCCGANGT 515  
OY 261 nAlaLeuGlYTyraAsnAspGlutThSerArghIsAlaLeuAArgYrArgrglutThVa 281  
DB 516 ATTGCTGGTCTGCGCGATACGAA-----GAACATCGCGATGATAT 557  
OY 281 lGlutInserrProLeuTyrglnlYsIleglThrclYarg 294  
DB 558 TGAACGCCATTATTTGAANAATCTCATCTCAGCGCGCGCG 597

RESULT 12  
LOCUS BM088538 549 bp mRNA linear EST 19-NOV-2001  
DEFINITION 502040 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BM088538  
VERSION BM088538.1 GI:16999166  
KEYWORDS EST.  
SOURCE  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
1 (bases 1 to 549)  
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Casas, E., Wray, J.E., White, J., Cho, J., Fahrnerkrug, S.C., Bennett,  
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,  
Pettea, G., Holt, I., Karameycheva, S., Liang, F., Quakenbush, J., and  
Keeler, J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)  
21180013  
Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@mealmarc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCATATGACCAT  
BACKWARD: GTTTCOCAGTCAGCAGC  
Plate: 1 row: I column: 7  
Seq primer: ATTTAGGTGACACTATGAC.  
Location/Qualifiers  
1..549  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
library made from pooled tissue from testis, thymus,  
semiteendonosus muscle, longissimus muscle, pancreas,

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
SOURCE



AU212487/c  
 LOCUS AU212487 704 bp mRNA linear EST 17-JUL-2001  
 DEFINITION AU212487 unpublished oligo-capped cDNA library, stage L4  
 Caenorhabditis elegans CDNA clone yk785f12 3', mRNA sequence.  
 ACCESSION AU212487  
 VERSION AU212487.1 GI:14850586  
 KEYWORDS EST.  
 SOURCE Caenorhabditis elegans.  
 ORGANISM Caenorhabditis elegans.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
 ; Rhabditidae; Pelodierinae; Caenorhabditis.  
 REFERENCE 1 (bases 1 to 704)  
 AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.  
 and Sugano,S.  
 TITLE A complementary view of the C.elegans genome  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Yuji Kohara  
 Genome Biology Lab.  
 National Institute of Genetics  
 Yata 1111, Mishima, Shizuoka 411, Japan  
 Tel.: 81-559-81-6854  
 Fax: 81-559-81-6855  
 Email: ykohara@lab.nig.ac.jp.  
 FEATURES  
 source  
 1. 704  
 /organism="Caenorhabditis elegans"  
 /strain="N2"  
 /db\_xref="taxon:6239"  
 /clone="yk785f12"  
 /clone\_lib="unpublished oligo-capped cDNA library, stage  
 L4"  
 /sex="hermaphrodite"  
 /tissue\_type="whole animal"  
 /dev\_stage="L4"  
 /note="The AD-wrmcDNA library was generated with poly(A)+  
 RNA isolated from both hermaphrodite and male N2 worms of  
 all larval stages, embryos, adults and dauers and the  
 subsequent generation of cDNAs by poly(A) priming. The  
 cDNAs were cloned into pPC66"  
 BASE COUNT 205 a 169 c 144 g 185 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.00971 Length: 704  
 Score: 115.00 Matches: 56  
 Percent Similarity: 39.90% Conservative: 23  
 Best Local Similarity: 28.28% Mismatches: 67  
 Query Match: 4.24% Indels: 52  
 DB: 9 Gaps: 11  
 US-10-066-551-4 (1-525) x AU212487 (1-704)  
 QY 79 LeuSnAspProHsGluAlaPheAlaAlaAlaAlaLeu-----IleGluSerAla 96  
 Db 700 CTCACACACCCAGCCGCTACATGGAGCTCGACGCAATTGCTCGAATTGACCCGCC 641  
 QY 97 GluHsSerLeuAspLeuGln-----TyrTyrIleTyrArgAsnAsp 110  
 Db 640 AAGAAATTCATTGATCATCTATGAGCTATTTCCGTTGTTTCATCTACCGTATCA 581  
 QY 111 IleSerGlyArgLeuLeuPhe-----AsnLeuValTyrLeuAlaIleGluArg 126  
 Db 580 -----CGTGTACATTTTCTCGAATATTGATGACGGATCCGCGGGGTGTAGA 530  
 QY 127 GlyValArgValArgLeuLeu-----LeuAspAspAsnAsnThrArgGly 141  
 Db 529 GGGCGCAAAATTCGATTCCTAGCGCCGCGCTCTCATTTCCGGAATTTGGAACGAGTTC 470  
 QY 142 LeuAspAspLeuLeuAlaLeuAspSerHisProAsn-----IleGluValArgLeu 159  
 Db 469 CTGAGATCTCTTGGAATCGTGAATGTTTCATGCTAATGACTATGAGTAAGTAAATTC 410  
 QY 160 PheAsnProPheValLeuArgLysTyrParGAlaLeuGlyTyrLeuThrAspPheProArg 179

Db 409 TTCAAA-----|||||  
 QY 180 LeuAsn-----|||  
 Db 394 ACCAACATTGAAGTATAGTAAATTAATCGGAGCGCGGACCATTAATAA--TTCATG 338  
 QY 191 AlaAspAsnArgAlaThrIleLeuGlyArgAsnIleGlyAspGluTyrPheLysVal 210  
 Db 337 GTCACCGAATCGCGCGGATTTATGGAACCTCAATTTGGTCGGAGACTATTTATGGT 278  
 QY 211 GlyGluAspThrValPheAlaAspLeuAspIleLeuAlaThrGlySer-----Val 227  
 Db 277 GGA-----ACTACTGAGCTCGCATAGTATACAGCAACCGCGGAGAGAGCGCTTTT 224  
 QY 228 ValGlyGluValSerHisAspPheAspArgTyrTyrPalaSerHisSerAlaHis 245  
 Db 223 GTGATGAGTTGAAGATATCTTACCCGCGAGTGGGAAAGTCATATTCAC 170  
 RESULT 15  
 D79520 570 bp mRNA linear EST 09-FEB-1996  
 LOCUS D79520/c  
 DEFINITION HUM272G08B Human aorta polyA+ (Tfujiwara) Homo sapiens CDNA clone  
 GEN-272G08 5', mRNA sequence.  
 ACCESSION D79520  
 VERSION D79520.1 GI:1179871  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 570)  
 AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,  
 Okuno,S., Ozaki,K., Shimizu,F., Shinoda,Y., Shinomiya,H., Takachi,  
 A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., Maekawa,H.,  
 Shin,S. and Nakamura,Y.  
 TITLE Fujiwara et al. (1995)  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Tsutomu Fujiwara  
 Otsuka GEN Research Institute  
 Otsuka Pharmaceutical Co.,Ltd  
 463-10 Kagasuno, Kawuchi-cho, Tokushima, Tokushima, 771-01 Japan  
 Tel.: 0886-65-2888  
 Fax: 0886-37-1035.  
 FEATURES  
 source  
 1. 570  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="GEN-272G08"  
 /clone\_lib="Human aorta polyA+ (Tfujiwara)"  
 /tissue\_type="aorta"  
 /note="Organ: brain"  
 BASE COUNT 123 a 169 c 143 g 135 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.00896 Length: 570  
 Score: 114.00 Matches: 48  
 Percent Similarity: 39.06% Conservative: 27  
 Best Local Similarity: 25.00% Mismatches: 84  
 Query Match: 4.20% Indels: 33  
 DB: 14 Gaps: 7  
 US-10-066-551-4 (1-525) x D79520 (1-570)  
 QY 102 LeuGlnTyrTyrIleTyrParGAsnAspIleSerGlyArgLeuLeuPheAsnLeuValTyr 121  
 Db 548 CTGACCGCTTATCTGTTGGGATGACGCGCAACACTGCATCGCGCACTACTG 489  
 QY 122 LeuAla-AlaGluArgGlyValArgValArgLeuLeuAspAspAsnAsnThrArgGly 141  
 Db 488 GCAGCAGCCCAACGCGGGGTTAAAGCGGAAGTCTTGCTGATGGCTACGGTCCGGA 429





QY 460 LysilealaglInmetcgu 466  
 DB 654 AACCTGGCTCAGACCTTGAG 674  
 RESULT 19  
 AA207840/c 548 bp mRNA linear EST 12-MAR-1997  
 LOCUS mw81b08.r1 guaywoodford Beiler mouse kidney day 7 Mus musculus cDNA  
 DEFINITION clone IMAGE:661431 5' similar to SM:U1P2.P31048 HYPOPHYSICAL  
 54.3 KD PROTEIN IN LPD-3 5' REGION ; mRNA sequence.  
 ACCESSION AA207840  
 VERSION AA207840.1 GI:1804371  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 548)  
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,K.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Maria M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:407279  
 Possible reversed clone: similarity on wrong strand  
 Seg primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 451.  
 FEATURES  
 source  
 1. 548  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:661431"  
 /clone\_lib="Guaywoodford Beiler mouse kidney day 7"  
 /tissue\_type="kidney"  
 /dev\_stage="juvenile (7 days old)"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: kidney; Vector: pBluescript SK-; Site\_1:  
 EcorI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
 oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;  
 -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor  
 sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' library provided  
 by Lisa Guay-Woodford."  
 BASE COUNT 122 a 151 c 157 g 118 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.025 Length: 548  
 Score: 110.00 Matches: 40  
 Percent Similarity: 33.67% Conservative: 27  
 Best Local Similarity: 20.10% Mismatches: 70  
 Query Match: 4.05% Indels: 62  
 DB: 9 Gaps: 5  
 US-10-066-551-4 (1-525) x AA207840 (1-548)  
 QY 335 LysSerValTyrLeuValSerProTyrPheValProThrLysSerGlyThrAspAlaLeu 354  
 DB 542 CCGGAGGTATGATCGCAAGCGCTACTCTTCCCGGCGATGATTTTACAGCGCTTG 483  
 QY 355 AlaLysLeuValGlnAspGlyIleAspValThrValLeuThrAsnSerLeuGlnAlaIaThr 374

DB 482 CGTAAGCGGACGCGCGGGGTGCGGATCAACTGATCATTCAGGCGC-----GAACCG 429  
 QY 375 AspValAlaAlaValHisSerGlyTyrValLysTyrArgLysProLeuLeuLysAlaGly 394  
 DB 428 GATATGCCGATTTGTACAGAGTCGGTGGCGCTTCGTATACATCATCTGGTTAAAGCGCGC 369  
 QY 395 IleLysLeuTyrGlnLeuGlnProAsnHisAlaValProIaThrLysAspLysGlyLeu 414  
 DB 368 GTTCAGGTTTGTAGTACCGCCGCCG----- 342  
 QY 415 ThrGlySerValThrSerLeuHisAlaLysThrPheIleValAspGlyLysArgIle 434  
 DB 341 -----CCGCTCCAGCGCAAAAGTGCGATGATGAGCATCACTGGCGC 300  
 QY 435 PheIleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThrGlnMetGlyVal 454  
 DB 299 ACAGTAGGGGTCAGTATCTCGATCGCTCAGTTGTCTACTGATCTCGAAGCAAAATGTC 240  
 QY 455 ValIleGlySer-----Pro 459  
 DB 239 ATCATCCAGCATGCTCATTTTAAACACGACTGGCGCATATCTGAACGCGATTATTGCCG 180  
 QY 460 LysIleAlaGlnInmetGluArgThrLeuAlaAspThrSerPro----- 474  
 DB 179 CAGATTGTCTACGAGTGATGAACCATGCTGCCCAACGACCTGGTGACCTGACCA 120  
 QY 474 ----- 474  
 DB 119 AAAGCTGCTGCGCTTCACCTTTTACAGCCACTTCCCGCGCTGGTGGCTTCGCG 60  
 QY 475 -----GluTyrAlaTyrArgValThrLeuAspArgHisAsnArgLeuGlnIlePHis 491  
 DB 59 CACACACGCCACGCTCGGCCAGGTGATGCCGCCGACACACGA---CATGTGAC 6  
 RESULT 20  
 BG920511 766 bp mRNA linear EST 05-JUN-2001  
 LOCUS 602824841F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:4953491 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG920511  
 VERSION BG920511.1 GI:14300987  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 766)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey Green M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM10913 row: a column: 12  
 High quality sequence stop: 766.  
 FEATURES  
 source  
 1. 766  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4953491"  
 /clone\_lib="NCI\_CGAP\_Mam6"  
 /sex="female, virgin"  
 /tissue\_type="infiltrating ductal carcinoma"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;







Y	107	TripharyngaspiliserGlyArpleu	-----AspleValTyrleu	122
	107			
	571	TACCGTAATCCA-----CGTGTACATTTCTCGAATATTGATGACCGATCCGACGG		521
	123	AlaIaIaGArGlyValArGValArGleu	-----LeuAspAspAsn	137
	111			
	520	GGCGTGGTAAAGCGCTCAAAATTCGATTCCTAGCGCGCGCTTCATTAACCGGATATT		461
	138	AsnThrArGlyLeuAspAspleuLeuAlaLeuAspSerHisProAsn-----Ile	155	
	460	GGAACGAGGTTCCGAGATCCCTTGAATCGTTGAATGCTTCATGCTAAAGGTACTATG	401	
	156	GlUValArGleuPheAsnProPheValLeuArGlyStrPrArGAlaLeuGlyTyrleuThr	175	
	400	GAACTGAAATCTTCANA-----	383	
	176	AspPheProArGleuAsn-----ArgArgMethHisAsn	186	
	382	---GTTCCACCTACCAACATTAAGATTAATTAATTCGCGAGCCGCGGACCGCATTAAT	326	
	187	LysSerPheThrAlaAspAsnArGAlaThrTlleuGlyGlyArGAsnIleGlyAsp	206	
	325	AAA---TTCATGTCACCCGAGTCGCGGGAATTTATGSAACCTCAAAATGTGCGGGAAC	269	
	207	TyrPheLysValGlyGlyAspThrValAlaPheAlaAspleuAspIleLeuAlaThrGlySer	226	
	268	TATTTATGGGTGA-----ACTACGTGACCTCGCATAGTATCAGGCANAACGGGGAG	215	
	227	-----ValValGlyGlyValSerHisAspPheAspArgTyrTPPrAlaSerHisSer	243	
	214	AAGCGCGCTTTTGTGATGAGTCAAGCATATCTTCAACCCGCGACTGGGAAAGTGCATAT	155	
	244	AlaHis	245	
	154	TCCAC	149	
RESULT 25				
LOCUS	BJ130081/c	649 bp	linear	EST 23-JAN-2002
DEFINITION	BJ130081 unpublished oligo-capped cDNA library, C. elegans LI stage			
ACCESSION	BJ130081			
VERSION	BJ130081.1	GI:18290238		
KEYWORDS	EST.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans.			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea			
AUTHORS	1 Rhabditidae; Pelodierinae; Caenorhabditis.			
	1 (phases 1 to 649)			
	Kohara,Y., Shih-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.			
	and Sugano,S.			
TITLE	A complementary view of the C.elegans genome			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Tadasu Shin-i			
	Center For Genetic Resource Information			
	National Institute of Genetics			
	1111 Yata, Mishima, Shizuoka 411-8540, Japan			
	Tel: 81-559-81-6856			
	Fax: 81-559-81-6855			
	Email: tshini@genes.nig.ac.jp.			
	Location/Qualifiers			
	1. 649			
FEATURES				
source	/organism="Caenorhabditis elegans"			
	/strain="N2"			
	/db_xref="taxon:6239"			
	/clone="YK1038e09"			
	/clone_lib="unpublished oligo-capped cDNA library, C.			
	elegans LI stage"			
	/sex="hermaphrodite"			
	/tissue_type="whole animal"			
	/dev_stage="L1"			
	/note="The AD-wtmcDNA library was generated with poly(A)+			

[illegible]

Db 619 AACCACTTCTATCTGGGACGCCCAACATGAGCTGGCTGCTGACGACGATGAAGAG 678  
 QY 452 MetgylvalVal 455  
 Db 679 CTGGGCGTGTCTC 690

RESULT 27  
 BE206365 955 bp mRNA linear EST 06-NOV-2000  
 LOCUS 601869921F1 NIH\_MGC\_19 Homo sapiens cDNA clone IMAGE:4100068 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BE206365  
 VERSION BE206365.1 GI:11099951  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 955)  
 AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at: image.lnl.gov  
 Plate: LICM69 row: 3 column: 05  
 High quality sequence stop: 640.

FEATURES  
 Source Location/Qualifiers  
 1..955  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4100068"  
 /clone\_1db="NIH\_MGC\_19"  
 /tissue\_type="neuroblastoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: brain; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Library constructed by Ling Hong  
 in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

BASE COUNT 224 a 297 c 279 g 155 t  
 ORIGIN

Alignment Scores:  
 Pired. No.: 0.109 Length: 955  
 Score: 108.00 Matches: 68  
 Percent Similarity: 41.63% Conservative: 39  
 Best Local Similarity: 26.46% Mismatches: 86  
 Query Match: 3.96% Indels: 65  
 DB: 12 Gaps: 15

US-10-066-551-4 (1-525) x BE206365 (1-955)

QY 125 GUAARGGLVALARGVALARGLEULEULEUASP-----ASpASnASrThArgGlyLeu 142  
 Db 163 GAGGCTGGCGTCAAGGTCCGCTCATCTGCTGGGAGACACTCGAGCCATCCATG 222  
 QY 143 AspAspLeuLeuAlaLeu-----AspSerHisProAsnIleLeu 156  
 Db 223 CGGGCCCTTCCTGCTCTCTGCTGCTGCGTGAACAACATACCACCTGATCCATCCAG 282  
 QY 157 ValArgLeuPheAsnProPheValLeuArgLysTrpArgAlaLeuGlyTyrLeuThrAsp 176  
 Db 283 GTGAACACTC-----TTTGTGTGTCCCGCGGATGAGGCCACGCTCGAATC---CCA 330

QY 177 PheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAspAsnArgAlaThr 196  
 Db 331 TATGCCCGCTGTCAAC-----CACAAAG---TACATGTGATGACGAGCGGCCACC 378  
 QY 197 IleLeuGlyArgAsnIleGlyAspGluTyrPheLeuValGlyGluAspThrValPhe 216  
 Db 379 TACATCGGAGACCTCCACTGTGCTGGCACTTTC-----ACGGAGACG 423  
 QY 217 AlaAspLeuAspIleLeuAlaThr-----GlySerValValGlyValSer 232  
 Db 424 CGGGCACCCTGCTGCTGTGTGATGACGAGATGAGGAGGGCGGCTGCGGACGACCTGAG 483  
 QY 233 HisAspPheAspArgTyrTrpAlaSerHisSerAlaHisAsnAlaThr----- 248  
 Db 484 GCCATTTTCTGAGGAGGATGGGAGCTCCCTTACAGCCATGACTGACACCTCAGCTGACA 543  
 QY 249 -----ArgIleIleArgSerGlyAsnIleGlyLys----- 258  
 Db 544 CGGTGGGAGACGCTGCTGTGAGGCGCCGATCCAGTGGGAGGAGGAGGCTGTGCTGGG 603  
 QY 259 -----GlyLeuGlnAlaLeuGlyTyrAsnAspGluThrSerArgH 272  
 Db 604 CCGGGGACCAAGTGTCTGTGGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 663  
 QY 272 SalLeuLeuArg-----TyrArgGluThrValGluGlnSe 284  
 Db 664 TGTGCTCTCAGATCTTACATGTGACGACCTATAGATACACACACGCGTACGCGCG 723  
 QY 284 rProLeuTyrGlnLysIleGlnThrGlyArgIleAspTrpGln-----SerVa 300  
 Db 724 CGCCGG-ACCAAGAGAGCTG-----GGAGAGATAGACCAAGAGGCGGTAGACAGAG 776  
 QY 300 IgLThrArgLeuIleSerAspAspProAlaLysGlyLeuAspArgAspArgLysPr 320  
 Db 777 GCAAAAC---ATACAGGAGAGAAAGCGCCAGCGGAGATCGGGGAGCAGCAACAC 833  
 QY 320 opIroIleLaGlyArgLeuGlnAspAlaLeuLysGlnProLysSer 336  
 Db 834 CCAAGCAAAACGAGAGACCGA-----AGACAACCCGAGAGAAC 873

# RESULT 28

AM437576 440 bp mRNA linear EST 25-APR-2001

LOCUS 78841 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.

DEFINITION AM437576

ACCESSION AM437576.1 GI:6972882

VERSION EST.

KEYWORDS

SOURCE

ORGANISM

Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;

Bovidae; Bovinae; Bos.

1 (bases 1 to 440)

Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,

Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,

G.L., Heaton,M.P., Leigred,W.W., Rohrer,G.A., Chitko,W.K., C.G.,

Pertea,J.G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and

Keefe,J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA

libraries and construction of a gene index for cattle

Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith TPL

USA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 20

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

BACKWARD: GTTTCCAGTCACGACG  
 Plate: 42 row: L column: 11  
 Seq primer: ATTGATGACATCTATAG.

## FEATURES

Location/Qualifiers  
 1..440

## SOURCE

/organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="MARC 1B0V"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"

/note="Vector: PCMV SPOR6; Site\_1: NotI; Site\_2: SalI;  
 library made from pooled tissue from lymph node, ovary,  
 fat, hypothalamus, and pituitary."

## BASE COUNT

85 a 132 c 144 g 79 t

## ORIGIN

## Alignment Scores:

Pred. No.: 0.0343 Length: 440  
 Score: 107.50 Matches: 36  
 Percent Similarity: 50.50% Conservative: 15  
 Best Local Similarity: 35.64% Mismatches: 35  
 Query Match: 3.96% Indels: 15  
 DB: 10 Gaps: 5

US-10-066-551-4 (1-525) x AW437576 (1-440)

QY 369 AsnSerLeuGlnAlaThrAspValAlaValAlaHisSerGlyTyrValLysTyrArgLys 388

DB 42 AACGGCCTGCAGACT-----GACGGCCTGGCTGTGTGGCAGAC 80

QY 389 ProLeuLeuLysAlaGlyLeuTyrGluLeuGlnProAsnHisAlaValProAla 408

DB 81 CAAGATGACAGCCAGACGGCGGACCTTGAGAGTCCAGCTGAGACAC--GTGCCCATG 137

QY 409 ThrLysAspLysGlyLeuThrGlySerSerValThrSerLeuHisAlaLysThrPheLe 428

DB 138 GGAAAG-----CTCAGTGGCGGC-----GTTTTCATCCCAAGTTCTGGGTG 179

QY 429 ValAspGlyLysArgLeuPheLeuIleGlySerPheAsnLeuAspProArgSerAlaArgLeu 448

DB 180 GTGGATGGCGGACATCTACGTGGGACGTGCCACATGACATGGCGGTCTCCGACACAG 239

QY 449 AsnThrGluMetGlyValIleLeuGluSerProLysIleAlaGluGlnMetGluArg 467

DB 240 GTGACGAGCTGCTGCTGCTGATCATCTACACTGACGACCGCTGCGCCAGACCTCGAAG 299

QY 468 Thr 468

DB 300 ACC 302

## RESULT 29

BI522691 912 bp mRNA linear EST 28-AUG-2001

DEFINITION 603175730F1 NIH\_MGC\_121 Homo sapiens cDNA clone IMAGE:5240191 5',

ACCESSION BI522691

VERSION BI522691.1 GI:15347483

KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgi.nci.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC).

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cga@bbs-remail.nih.gov

Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLML1605 row: k column: 08  
 High quality sequence stop: 898.

## FEATURES

## SOURCE

1..912  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:5240191"  
 /clone\_lib="NIH\_MGC\_121"  
 /lab\_host="DH10B"

/note="Organ: brain; Vector: PCMV-SPOR6; Site\_1: NotI;  
 Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3  
 fetal brains, female age 20 weeks, female age 24 weeks,  
 and male age 26 weeks. Library is oligo-dT primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.7 kb, insert size range  
 0.7-3.5 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 017. Note:  
 this is a NIH-MGC Library"

## BASE COUNT

188 a 288 c 270 g 165 t

## ORIGIN

## Alignment Scores:

Pred. No.: 0.2 Length: 912  
 Score: 105.50 Matches: 52  
 Percent Similarity: 38.57% Conservative: 29  
 Best Local Similarity: 24.76% Mismatches: 69  
 Query Match: 3.89% Indels: 60  
 DB: 13 Gaps: 10

US-10-066-551-4 (1-525) x BI522691 (1-912)

QY 319 LysProProIleAlaGlyArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyr 338

DB 321 CACCCAGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGT 380

QY 339 LeuValSerProIle-----PheValProThrLys--- 348

DB 381 CTACTGACCTTACACACATAC 440

QY 349 SerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAspValThrValLeuThr 368

DB 441 GAGAGTCTCTCCGACAGTGCAGACCTGTGGACCAAGAGGGGTGACGTCCGACATCGCTGT 500

QY 368 PAsn-----SerLeuGlnAlaThrAspValAlaValHisSerGlyTyr 383

DB 501 GACCAAGCCCGACCGGCGCCAGGACGACGCTGTGCTGTGACGAGCGGT-- 558

QY 383 ValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnPro-- 402

DB 559 -----GCCAGGTCGCGATGCTGACATGACATGCAAGCT 590

QY 403 AsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerLe 422

DB 591 GACCCATGGCGTC-----CT 605

QY 422 ValHisAlaLysThrPheIleValAspGlyLysArgIlePheIleGlySerPheAsnLeuAs 442

DB 606 GCATACCAAGTCTGTGGTGTGGACAGACCCACTTCTGAGGACGATGCCAACATGGA 665

QY 442 pProArgSerAlaArgLeuAsnThrGluMetGlyValIleLeu----- 466

DB 666 CTGGCGTCTACCTGACCCAGCTCAAGAGCGTGGCTGATGTATACAACTGACAGCTGCCCT 725

QY 457 -----GluSerProLysIleAlaGlu-----GluMetGluArgThr 468

DB 726 GGCTCGAGACCTGACAAAGATCTTTAGAGGCTTACTGCTTCTGGCCAGGACGACGCTC 785

QY 468 rLeuAlaAspThrSerProGluTyr-----AlaTyrArgValThrLeuAspArgH 485

DB 786 CATCGCATCAACTTGGCCCGGTTTATATGACACACCGCTACAAACAGACACCAATGGA 845

**OY**     **485**   sasnarlgcuglnrprhisasprrcola 494  
           ||||| ||||| ||||| :::|||||  
**Db**       **846**   GAATCTGCTCAATG---AACCTTGCT 870  
  
**RESULT 30**  
**LOCUS**       BE332913  
**DEFINITION** BE332913 632 bp mRNA linear EST 14-JUL-2000  
                 usg6a10.y1 Soares thymus 2NbMv Mus musculus cDNA clone  
                 IMAGE:3326202 5' similar to TR:Q92853 Q92853 HU-KA. :, mRNA  
                 sequence.  
**ACCESSION**   BE332913  
**VERSION**     BE332913  
**KEYWORDS**   EST.  
**SOURCE**      house mouse.  
                 Mus musculus  
**ORGANISM**   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
                 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**REFERENCE**   1 (bases 1 to 632)  
                 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**TITLE**       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
                 Tumor Gene Index  
**JOURNAL**     Unpublished (1997)  
**COMMENT**      Contact: Robert Strausberg, Ph.D.  
                 Email: cgapbs-r@mail.nih.gov  
                 This clone is available royalty-free through LINT ; contact the  
                 IMAGE Consortium (info@image.llnl.gov) for further information.  
                 MG1:1070366  
                 Seq primer: -40RP from Gibco  
                 High quality sequence stop: 467.  
                 Location/Qualifiers  
                     1..632  
                     /organism="Mus musculus"  
                     /strain="C57BL/6J"  
                     /db\_xref="taxon:10090"  
                     /clone IMAGE:3326202"  
                     /clone\_id="Soares\_thymus\_2NbMv"  
                     /sex="male"  
                     /tissue\_type="Thymus"  
                     /dev\_stage="4 weeks"  
                     /lab\_host="DH10B"  
                     /note="Vector: pRT73D-Pac (Pharmacia) with a modified  
                         polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
                         was primed with a Not I - oligo(dT) primer [5'  
                         TGTTCACCATCTGAGTAGGGAGCGCCGCCGCTTTTGTATTGTTTTTTTTTTTTTT  
                         3'] double-stranded cDNA was ligated to Eco RI adaptors  
                         (Pharmacia), digested with Not I and cloned into the Not I  
                         and Eco RI sites of the modified pRT73 vector. RNA  
                         provided by Dr. Bertrand Jordan. Library went through two  
                         rounds of normalization, and was constructed by Bento  
                         Soares and M.Fatima Bonaldo."  
**BASE COUNT**   125 a       201 c       158 g       147 t       1 others  
**ORIGIN**  
  
**Alignment Scores:**  

Pred. No.:	0.124	Length:	632
Score:	105.00	Matches:	35
Percent Similarity:	45.61%	Conservative:	17
Best Local Similarity:	30.70%	Mismatches:	30
Query Match:	3.87%	Indels:	32
DB:	10	Gaps:	6

  
**US-10-066-551-4 (1-525) x BE332913 (1-632)**  
**OY**     **362**   ileaspaValthVaiLeuthrasnsr-----LeuglAlaIathrasp----- 375  
           |||   |||   |||   |||||:  
**Db**     **15**   ATCTCTGTGGTGIGTGGCCACCACGCCACACATTGGCCACAGACATCCACTGACTCCAG 74  
  
**OY**     **376**   valalaalavalnisserclytyValyalsyltyrarglsProLeuDeuLyslaaglytle 395  
           |||   |||   |||   :::  
**Db**     **75**   GCCTTGCGTGGCCAGTGGTGGCCAGATFAGA----- 104  
  
**OY**     **386**   lysleutyrGluleuGlInPrOAsnhIsalayaValProAlaIathrLysaspLysglyleuthr 415

Qy	Db	105	-----	CAAGTGGCCATG-----					AAACAGCTTACT	128
Qy	416	GlySerSerValThSerLeuHisAlaIylStrPrheIIleValAspGlyLysArgIlePhe	435							
Db	129	GGGGGT-----	GTTCACACACGCCAAATTCCTGGTTGTGATGGGGGACACACTAC	179						
Qy	436	IleGlySerPheAsnLeuAspProArgSerAlaArgLeuAsnThGlnMetGlyValVal	455							
Db	180	GTGGCACAATGCCAACATGAGCTGGCGCTCCCGACTCAGCTCAAGGAACCTTGTCGCAAC	239							
Qy	456	Ile---GluSerProLysIleAlaGluGlnMetGluArgThr	468							
Db	240	ATCTCAACACTGCAGCAACCTGGCTCAAGACCTTGAGAAACA	281							
RESULT	31									
LOCUS	AM030783		366 bp	mRNA	linear	EST 18-MAY-2001				
DEFINITION	EST274038 tomato callus, TAMU Lycopersicon esculentum cDNA clone									
ACCESSION	AM030783									
VERSION	AM030783.1		GI:5889539							
KEYWORDS	EST.									
SOURCE	tomato.									
ORGANISM	Lycopersicon esculentum									
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.									
AUTHORS	1 (bases 1 to 366) Alcala, F., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E., Liang, F., Upton, D., Craven, M.B., Bowman, C.L., Ann, S., Rönning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, D.									
TITLE	Generation of ESTs from tomato callus tissue									
JOURNAL	Unpublished (1999)									
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: <a href="http://www.genome.clemson.edu/orders/index.html">http://www.genome.clemson.edu/orders/index.html</a> 5 prime sequence.									
FEATURES	Location/Qualifiers									
source	1..366 /organism="Lycopersicon esculentum" /cultivar="TA96" /db_xref="taxon:4081" /clone="CLOC22114" /clone_1db="tomato callus, TAMU" /tissue_type="callus" /dev_stage="25-40 days old" /lab_host="X1L-Blue MRF" /note="Vector: pBluescript SK(-); Site_1: EcoRI, Site_2: XhoI; supplier: Giovannoni Laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato callus EST library"									
BASE COUNT	86 a	85 c	107 g	88 t						
ORIGIN										
Alignment Scores:										
Pred. NO.:	0.0572	Length:	366							
Score:	104.50	Matches:	33							
Percent Similarity:	46.15%	Conservative:	21							
Best Local Similarity:	28.21%	Mismatches:	56							
Query Match:	3.85%	Indels:	7							
DB:	10	Gaps:	2							
US-10-066-551-4 (1-525) x AM030783 (1-366)										
Qy	76	IleYrIleuLeuAsnAspProHisGluAlaPheAlaIaArgAlaAlaLeuIleGluSer	95							
Db	35	ATTCAGTTGGCGAAGACGGCGACGACAAATTAATTCGCCGGCTGTTAAAGCGCATTTGGCGAG	94							

Oy	96	ALGIuHtISserLeuAspLeuGlnGlyTyrTrieTpaTgAsnAspILeserGlyArgLeu	115
		: : : : :	
	95	GCACAGAACCGCATTCCTTGAAACGGTTTATCTGTTGAGAGATGACGTCGGCACAACA	154
Db	116	LeuPheasLeuValTyrLeuAlaIaGluArgGlyValArgValArgLeuLeuLeuAsp	135
Oy	155	CTGCATCGCGGACACTCTGGACGACGGCAGCGGGGTAAAGCGGAAGTCTCGTGAT	214
Db	136	AspAsnAsnThrArgGlyLeuAspAspLeuLeuLeuAlaLeuAspSerHisProAsnIle	155
Oy	215	GGCTACGGTTCGCCGAGATCTCAGCGATGAGTTTGTCAAATGAACCTAGCGGACCTGGCGTA	274
Db	156	GIuValArgLeuPheasProPheValLeuArgGlySTPArgAlaLeuGlnGlyTyrLeuThr	175
Oy	275	GTGTCCCGCTACTACGATCC-----CGCCCTCGCCTTTGGTATGCGCAC	322
Db	176	AspPheProArgLeuAsnArgArgMetHisAsnLysSerPheThrAlaAsp	192
Oy	323	AAT-----GTGTTTCGGCGAGATGATGCCAANAATTGTGTATGTCAC	364
Db			

RESULT 32	LOCUS	DEFINITION
AU218786/c	628 bp	EST 17-JUL-2001
AU218786	unpublished oligo-capped cDNA library, stage L1	
AU218786	<i>Caenorhabditis elegans</i> cDNA clone yk869c03 3', mRNA sequence.	

ACCESSION AU218786 GI:14856943  
VERSION AU218786.1  
KEYWORDS EST.  
SOURCE Caenorhabditis elegans.  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae  
; Rhabditidae; Felodierinae; Caenorhabditis.

REFERENCE	1 (bases 1 to 628)
AUTHORS	Kohara,Y., Shin-I,T., Thierry-Mieg,D., Suzuki,Y. and Sugano,S.
TITLE	A complementary view of the C.elegans genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Yuji Kohara

Genome Biology Lab.  
National Institute of Genetics  
Yata 111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855  
Email: ykohara@lab.nig.ac.jp.  
Location/Qualifiers  
1. .628

```

/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone_id="YK869C03"
/clone_lib="unpublished oligo-capped cDNA library, stage
L1"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L1"
/notes="The AD-wrmcDNA library was generated with poly(A)+
RNA isolated from both hermaphrodite and male N2 worms of
all larval stages, embryos, adults and dauers and the
subsequent generation of cDNAs by poly(A) priming. The
cDNAs were cloned into pPC86"
BASE COUNT      164 a      151 c      123 g      168 t      2 others
ORIGIN

```

Alignment Scores:	
Pred. No.:	0.141
Score:	104.50
Percent Similarity:	38.95%
Best Local Similarity:	29.65%
Query Match:	3.85%
DB:	9
Length:	628
Matches:	51
Conservative:	16
Mismatches:	60
Indels:	45
Gaps:	10

US-10-066-551-4 (1-525) x AU218786 (1-628)

QY	98	HisSerLeuAsp---	LeuclhlyrlyrllrletrpagsAspIleSerlylrargleu	116
Db	609	CATGTTATGACATATTTCCCGTGTTCATCTACCGGTAATCCA-----	CGTGCACAT	559
QY	117	Phe-----	AsnleuValTyrleuAlaAlaIguArglylValArglyl	132
Db	558	TTCTCGAATATGTGATGACGCGCATCGACGGCGGTGGTTAGAGCGGTCGTAATTTGCATTC		499
QY	133	Leu-----	LeuaspaspasnaInrTrpGlyLeuaspaspaleu	147
Db	498	CTAGCGCGCGCGCTTCATTTATCCGGATATTGGAAAGCAAGTCTCTCGATGCTTGGCAATCG		439
QY	148	AlaLeuaspSerhizProasn-----	IleAluValArglylLeuPheaspnProPheValLeu	165
Db	438	TTGATGTGTTTCCATGCTCATAGTACTAGTAAGAGTAAGAAATCTTCAAA-----		391
QY	166	ArglylstrpArgAlaLeuGlyTyrleuThrAspPheProArgLeuasn-----		181
Db	390	-----	GTTCACAGTACCAACACTTGAAGGTATA	364

```

0y 182 -----ArglrqymethIsanlsySserPheThrAlaAspmrIgrAlaThr 196
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 GTTAATTATTCGGAGACGCCGAGCATATTATAA---TTCAATGTCACCGAGTCGGCGGGA 307
0y 197 lIleuuglyglrYargAsnllleclYaspsglurYrphelYsvalGlyclunspThrValPhe 216

```

Db	306	ATTATTTGGACCTCAAAATTGCTGGGAGCTATTTATGGTGA-----ACTACTGGA	253
Qy	217	AlaAspIeuAspIleIeuLaIarhGlySer-----ValValGlyIuValSerHis	233
Db	252	GCTGGCATAGTTATACAGCAACCGGGGAGAACCGGCTTTTGTCATGATGATTGAAGCAT	193

DB	192	ATCTTCACCCGCGACTGGGAAGTGCATATPTCCAC	157
QY	234	ASPPHEASPAIRGYITRPALASERTHSERIALAHIS	245
RESULT 33			
BF152678/c			

LOCUS	Bj15267.8	/30 bp	mRNA	linear	Est. 24 JAN-2000
DEFINITION	Bj15267.8, unpublished oligo-capped cDNA library, C. elegans 41 stages				
ACCESSION	Caenorhabditis elegans cDNA clone	X1310d02.3			mRNA sequence.
VERSION	Bj15267.8	1			GI:18320663
KEYWORDS	EST.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Caenorhabditis elegans				

REFERENCE	Eukariota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS	1 (bases 1 to 730) Kohara, Y., Shin-I, T., Thliery-Mieg, J., Thliery-Mieg, D., Suzuki, Y., and Sugano, S.
TITLE	A complementary view of the C. elegans genome
JOURNAL	Unpublished (2002)
COMMENT	Contact: Tadasu Shin-I Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshin@genes.nig.ac.jp.
FEATURES	location:genes.nig.ac.jp.
source	1..730

```

/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="Yk1310d02"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"
/sex="hermaphrodite"
/tissue_type="whole animal"

```

```
/dev_stage="L1"  
/note="The AD-wrmcDNA library was generated with poly(A)+
```





```

Db 360 GTGCGACTCTTTCTGTTCCAAATCGCCGTCGTGCAATTTGG-----CTCGTG 313
QY 175 ThrAspPheProArgLeuAsnArgArgMet-----HisAsnLysSerPheThrAla 191
Db 312 GCCCAATTCGGCGCGCTCGTGAGCGCGCTGCGGTACACACACAAAATATTCCTCTTC 253
QY 192 AspAsnArgAlaThrIleLeuGlyGlyArgAsnIleGlyAspGluTyrPheLysValGly 211
Db 252 GACAACTGGCAGACGATTCCTACCGGGGGGCAACCTTCGAGCAGCATTTTCGACACCCGT 193
QY 212 GluAspThrValPheAla 217
Db 192 GTTGACCGCTATTGGCT 175
RESULT 35
LOCUS B1856614 1333 bp mRNA linear EST 10-OCT-2001
DEFINITION 603385754P1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5394589 5',
mRNA sequence.
ACCESSION B1856614
VERSION B1856614.1 GI:15997361
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/
1 (bases 1 to 1333)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLINL at:
http://image.llnl.gov
Plate: L1AM12005 row: 1 column: 14
High quality sequence start: 7
High quality sequence stop: 269.
Location/Qualifiers
source 1. 1333
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5394589"
/clone_lib="NIH_MGC_87"
/issue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: breast; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 319 a 524 c 305 g 185 t
ORIGIN
Alignment Scores:
Pred. No.: 0.652 Length: 1333
Score: 103.50 Matches: 81
Percent Similarity: 33.17% Conservative: 56
Best Local Similarity: 19.61% Mismatches: 149
Query Match: 3.81% Indels: 128
Gaps: 16
DB: 13
US-10-066-551-4 (1-525) x B1856614 (1-1333)
QY 171 LeuGlyTyrIleuThrAspPheProArgLeuAsnArgArgMetHisAsnLysSer----- 188
Db 173 GTGGGACAGAGCTCTTCTTCCAAAATATGAGGAGCTCTTTGTACCATGGAAGCCATGT 232

```

```

QY 189 -----PheThrAlaAspAsn-----ArgAlaThrIleLeuGlyArgAsnIle 203
Db 233 TTCTGATGCTTACAGTCTGGGAATGTGATGGCATCTGCTGACCTCAGTCAACAGCAG 292
QY 204 GlyAspGluTyrPheLysValGlyGluAspThrValPheAlaAspLeuAspIleLeuAla 223
Db 293 GGTGACGGG-----CCGGTTCTTACCGATACAGCACTGGCTGGATGATTTGGCTGGCC 346
QY 224 ThrGlySerValValGlyGluValSerHisAspPheAspArgTyrTrpAlaSerHisSer 243
Db 347 ACAGAGCAGTCTGCTCATGATGATCCAAATGG-----TCTGCTTCCAA 391
QY 244 AlaHisAsnAlaThrArgIleLeuArgSerGlyAsnIleGlyLysGly----- 259
Db 392 GCGCACAACTGGGACGCTCTCTAGTGAAGTGAAGACAGTGTCTGTGGACAGCGCGCAGCT 451
QY 259 ----- 259
Db 452 CTCGGCCGCCGCCCTTTTGGCTAGCGCCCTTCCAAATTCCTACAGTGTCTATG 511
QY 260 -----LeuGlnAla-LeuGly----- 264
Db 512 GGGCCCAAAAGTCCATCCGTGCGCTAGGCTACCCCAAGGCGCATGCTCGGGTGA 571
QY 265 -----TyrAsnAspGluThrSerArgHisAlaLeuLeuAr 276
Db 572 CCCCTCTTACGCCCGCCAGCAGCTAGTACAGGTAGTACGAGGACAGGCTCCCTGGCCA 631
QY 276 GTTATGAGGluThrValGluGlnSerProLeuTyrGlnLysIleGlnThrGlyArgIleAs 296
Db 632 CTATAGAGCCGCGAAGAGGCTCCACTCTTCACAG----- 668
QY 296 pTTPGlnSerValGlnThrArgLeuIleSerAspAspPro-----AlaLysGln 312
Db 669 -TGGAGACAGGCCCA-TCCACACTGCTGCTGTACACCAAGGCGGACAGCAGCGG 726
QY 312 Y-LeuAspArgAspArg-----ArgLysProIleLeuLag 324
Db 727 ACCTACACGCGCGCCCAAAACACACAGCCCTGCGGAGAGACAGACGCCCTCGGGCG 786
QY 324 LysArgLeuGlnAspAlaLeuLysGlnProGluLysSerValTyrLeuValSerProTyr 344
Db 787 CACCCCTTGGCAGCTGATCCCTGCGCCAGGCGCAGACAGACAGCGCTATGCGACATC 846
QY 344 heValProThrLysSerGlyThrAspAlaLeuAlaLysLeuValGlnAspGlyIleAsp 364
Db 847 ACCTGCCACAGCCGCCCCACAC-----CAACTGACTGTGGCA 885
QY 364 alThrValLeuThrAsnSerLeuGlnAlaThr-----AspValAlaAlaValHisSerG 382
Db 886 TCCAGAGAAACACACACACACCCCATGACATGAAMAAGTCAACCCATGGCGCATVACG 945
QY 382 LysTyrValLysTyrArgLysProLeuLeuLysAlaGlyIleLysLeuTyrGluLeuGlnP 402
Db 946 CATATCAACACAAACACCCCAACCAACGATCACCGGCG-----CACATCGGAC 993
QY 402 roAsnHisAlaValProAlaThrLysAspLysGlyLeuThrGlySerSerValThrSerL 422
Db 994 CAGCGCATGGACCGCGCCCAACAGAGACCCGAGAGCGCAACC-----A 1035
QY 422 euHisAlaLysThrPheIleValAspLysArgIlePheIleGlySerPheAsnLeuA 442
Db 1036 CACACACCGCGCCA-----CTAC 1053
QY 442 spProArgSerAlaArgLeuAsnThrGluMetGlyValIleGluSerProLysIleA 462
Db 1054 CCCCAACGCGCGCCAGCAGCCCTGCGCATCAACCTA-----ACGGGCAACACACA 1104
QY 462 laeGluMetGluArgThrLeuAlaAspThrSerProGluTyrAlaLys----- 478
Db 1105 CCTTACGAGCCCAAGTACACCTTAGCGACTGATCATCCCAACGACAGCAGCAGCAGCAGC 1164
QY 479 -----ArgValThrLeuAspArgHisAsnArgLeuGlnTrpHisAspProA 494

```



## ORIGIN

Alignment Scores:

Pred. No.:	0.185	length:	533
Score:	102.50	Matches:	46
Percent Similarity:	41.82%	Conservative:	23
Best local Similarity:	27.88%	Mismatches:	65
Query Match:	3.78%	Indels:	31
DB:	13	Gaps:	7

US-10-066-551-4 (1-525) x BJ326903 (1-533)

[illegible]

/organism="Dictyostelium discoideum"

```
/db_xref="taxon:44689"
```

```
/clone_lib="Dictyostelium discoideum cDNA library, AF"
```

```

/dev_stage=Aggregatio
a 87 c 114 q

```

0.209	Leng
-------	------

41.828	Conservative
07.000	Liberal

3.78%	Indels:
12	Cases:

## ORIGIN

Alignment Scores:

Pred. No.:	0.573
Score:	102.50
Percent Similarity:	41.82%
Best local Similarity:	27.88%
Query Match:	3.78%
DB:	13
Length:	573
Matches:	46
Conservative:	23
Mismatches:	65
Indels:	31
Gaps:	7

US-10-066-551-4 (1-525) x BJ328796 (1-573)

QY	321	ProteoliteaglygLeuGlnaspAlaLeuGlnaspGlnProGlyGlnSerValThyLeuVal	340
Db	73	CCATTAGGCTCTTAATTTACAAAGATGATCTGCACCTCATGATGCATGGATGAATTAAT	132
QY	341	SerProTyrrpheValProThrLysSerGlyThraspAlaLeuAlaLysLeuValGlnasp	360
Db	133	-----ACAAATGCA---ACGAATTCATTTGATATG	159
QY	361	GlyIleaspValThrValLeuThrAsnSerLeuGlnAlaThrAspValAlaAlaValHis	380
Db	160	GGTATTTTATTTATGACATTTAAACAGATGCGGCTCAATTAAGATCCAGTTTATGGTGACAG	219
QY	381	SerGlyTyrrValLysTyrrGlySerProLeuLeuLysAla-----GlyIle	395
Db	220	TTAGGTTTGGATATTAT---AAACCATTAAGTTGATGCAATTCAGACGGCTTGATATT	276
QY	396	LysLeuTyrrGluLeuGlnProAsnHisAlaValProAlaThrLysAspLysGlyLeuThr	415
Db	277	AGAAATGTACAAACATCAATCATCATCTATGCCAGATACCGATACGATCAAAATTTAGCT	336
QY	416	-----GlySerSerValThrSer-----Leu	422
Db	337	AAATGAGCTGTACAGATGATGATCAATCAATGGCCATCATTAAGTAGTGACAGTATTGTA	396
QY	423	HisAlaLysThrPheIleValAlaspGlyLysArgIlePheIleGlySerPheAsnLeuasp	442
Db	337	CACACCAAGGTTATGTGATGGATGAGTTAGCTGCTTATTTGGGTAGTGCAAAATTTAGAT	456
QY	443	ProArgSerAlaArgLeuAsnThrGlnMetGlyAlaValIleGluSer---ProLysIle	461
Db	457	TGGCGCTTCACCTGCACCAAGTTAAAGAATTTGGGTGTGCTATTCCAAAACGTCCATCAATG	516
QY	462	AlaGluGlnMetGlu	466
Db	517	GTATCAGATACCGAG	531
RESULT 39			
LOCUS	BQ387515	607 bp	mRNA linear EST 22-MAY-2002
DEFINITION	NISC_umm24c02.y1 NICHD XGC OVI Xenopus laevis cDNA clone		
ACCESSION	BQ387515		
VERSION	BQ387515.1	GI:21075202	
KEYWORDS	EST.		
SOURCE	African clawed frog.		
ORGANISM	Xenopus laevis		
REFERENCE	Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;		
TITLE	Xenopodinae; Xenopus.		
	1 (bases 1 to 607)		
	NH-XGC <a href="http://image.lnrl.gov/image/html/xenopuslib_info.shtml">http://image.lnrl.gov/image/html/xenopuslib_info.shtml</a> .		
	National Institute of Child Health and Human Development, National		
	Cancer Institute, Xenopus Gene Collection		

JOURNAL  
COMMENT

Unpublished (2002)  
Contact: Robert Strausberg, Ph.D.  
Email: c9abs@email.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov  
Plate: L14M1199 row: F column: 4  
Seq primer: M13RP1 reverse primer (ABI).

FEATURES  
SOURCE

Location/Qualifiers  
1. 607  
/organism="Xenopus laevis"  
/db\_xref="taxon:8355"  
/clone="IMAGE:5074179"  
/clone\_lib="NICHD XCG Ovl"  
/sex="female"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: PCMV-SPORT6; site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2.0 kb. Constructed by Life  
Technologies."

BASE COUNT 141 a 163 c 163 g 140 t  
ORIGIN

## Alignment Scores:

Pred. No.: 0.23 Length: 607  
Score: 102.50 Matches: 41  
Percent Similarity: 42.17% Conservative: 29  
Best Local Similarity: 24.70% Mismatches: 55  
Query Match: 3.78% Indels: 41  
DB: 14 Gaps: 8

US-10-066-551-4 (1-525) x BQ387515 (1-607)

```

OY 380 Hissergltgrrvallystrarglyserproleuleuysalaglyllelyseuylrclu 399
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2 CACGCGTCGGGAGCAGACATAC-----TTGTCAGTGTGTGGCTTCGTACCCACGGG 52
OY 400 Leuglnproasnhsalavalproalathrlyasprlysglyleuthrlyserseval 419
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 53 GACATGCCCGAT-----GCGCTCTCTCTT 76
OY 420 Thrsrleu-----HisAlalysThrphellevaalspglylsargllephelle 436
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 77 ACGGACACTGTTCATTCATTCAGCAGAGATGCTTATCGTGACAGAGATATATCAT 136
OY 437 Glyserpheasnleuasproariser-----Alaargleuasnthrlumetglyval 454
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 137 GGGTCGGCCACATCATGACAGAGCATGCTGGGAAACGAGATAGTGCCTGCTGTG 196
OY 455 Valileguserprolysilealaglulmetgluarqthrleualaspthrserpro 474
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 197 CTGCTGGAAGACAGAGATTGTGAGCATCATGATGATGACAGCCCTACAGCGGGG 256
OY 475 GlutyrAlatyrarvalThrleuasparthrhisasnarg----- 487
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 257 AAATTCCTCCACCTAGTAATGAGCTTTCATATACATTCTAGGGCAGCGGCTTC 316
OY 488 -----Leuglntrphlsasprroalathrarglystrlytrproasnsluprogluala 505
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 317 CCGTACACTGATGTCTGTCTGACAGCTTACGACCATTTCTTT-----AAGAG----- 364
OY 506 Lysleutrrlylsargllealala-----Lyslleuaserleu 518
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 365 -----GTTTGGAAACGACAGCTGCTACCAACACCAATATCTATGACCAAGATTTTCCGGTGC 421
OY 519 Leuprolleugluserleu 524
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 422 CTCGCCGCTATATGCAGTG 439

```

RESULT 40  
LOCUS

Bj282990 685 bp mRNA linear EST 09-APR-2002  
Bj282990 Y. Ogihara unpublished cdna library, wh\_r Triticum

DEFINITION  
aestivum cdna clone wh13910 3', mRNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE

Bj282990  
Bj282990.1 GI:20103997  
EST.  
Triticum aestivum  
bred wheat.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 685)  
Ogihara, Y. and Murali, K.  
Expressed genes in Triticum aestivum  
Unpublished (2002)  
Contact: Tadasu Shin-1  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp.

FEATURES  
SOURCE

Location/Qualifiers  
1. 685  
/organism="Triticum aestivum"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="wh13910"  
/clone\_lib="y" Ogihara unpublished cdna library, wh\_r"  
/tissue\_type="root"  
/dev\_stage="Feekes' scale 1"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
site\_1: EcoRI; site\_2: XhoI; Plants were grown under  
hydroponic conditions at UC Davis, salt stressed for 12  
hours, and for 7 days, then dissected and frozen (Akhunov  
in J. Dvorak Lab). Total RNA was prepared from sheath  
tissue, equal quantities of RNA were pooled from the two  
samples, polyA was purified from the pooled RNA, a cDNA  
library was made, and the cDNA clones were in vivo  
excised to give plasmid phagemids in the T7 close lab  
at the University of California, Riverside (Akhunov, Chin  
, Choi, Close, Fenton, Kianin, Otto, Simons, Zhang).  
Plasmid DNA preparations and DNA sequencing were  
performed in the OD Anderson lab (all other authors)."

BASE COUNT  
ORIGIN

Alignment Scores:  
Pred. No.: 0.281 Length: 685  
Score: 102.50 Matches: 54  
Percent Similarity: 40.98% Conservative: 30  
Best Local Similarity: 26.34% Mismatches: 80  
Query Match: 3.78% Indels: 41  
DB: 13 Gaps: 11

US-10-066-551-4 (1-525) x Bj282990 (1-685)

```

OY 206 Glutyrphelysvalgllyslaspthrvalpheliaspleuaspile-----leu 222
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 683 GAGCACTAAATAATAGCT--CTTACGGTGTGTGATGTTCTCTGTGATACGAACCTT 627
OY 223 Alathrglyservalvalgllyslaspthrhisaspleasp-----Argtyrrpala 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 626 ATGACAGGGGAATATTTGGCCCATGCTTCCCAATATTAACGGTCGACAAAGATTGAAGA 567
OY 241 SerhisseralalhisasnalaThrarglleleargserglyasnllleglylsglyleu 260
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 566 AAGCATCACACATCAAA-----CGCGGGGCGCAAGCCACTGCGGCGCTA 522
OY 261 Glinalaenuglytyraasnargluthrser----- 270
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: May 19, 2003, 18:16:47  
Job time : 1450 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:49:16 ; Search time 79 Seconds  
(without alignments)  
885.526 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 2713  
Sequence: 1 MRANKTOAMPSETISLMKT.....KIMKRIAKILSLPIESL 525

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2629	96.9	508	21	AAV75753
2	2599	95.8	508	21	AAV75752
3	2577.5	95.0	507	21	AAV75751
4	632	23.3	502	18	AAW55249
5	632	23.3	502	18	AAW55550
6	591	21.8	428	19	AAW98378
7	311	11.5	504	23	ABBA9813
8	295	10.9	495	23	ABP39428
9	293.5	10.8	480	23	ABBS4276
10	288.5	10.6	482	23	ABBA9993

11	279.5	10.3	206	18	AAW55249
12	275	10.1	493	23	ABBS32481
13	275	10.1	502	23	ABBS32493
14	264.5	9.7	494	23	ABP39331
15	243.5	9.0	422	22	ABP76548
16	243.5	9.0	500	23	AAW92760
17	243.5	9.0	500	23	AAW92760
18	238.5	8.8	442	22	AAW54491
19	237	8.7	442	22	AAW54491
20	180	6.6	356	22	AAW51963
21	176	6.5	280	22	AAW51152
22	176	6.5	1097	22	ABG01151
23	165	6.1	525	23	ABP28514
24	160.5	5.9	531	23	ABP28514
25	160.5	5.9	531	23	ABP28514
26	145.5	5.4	933	22	AAW70251
27	142.5	5.3	476	23	AAW50563
28	137	5.0	540	23	AAW16977
29	136	5.0	922	22	AAW70252
30	133.5	4.9	933	21	AAW44716
31	133.5	4.9	933	22	AAW5980
32	130.5	4.8	489	20	AAW36906
33	130.5	4.8	932	19	AAW53281
34	121.5	4.5	506	22	AAW73377
35	121.5	4.5	933	20	AAW89200
36	121.5	4.5	933	21	AAW44715
37	121.5	4.5	933	22	AAW85975
38	118	4.3	507	23	AAO17567
39	115	4.2	832	23	ABW2080
40	114.5	4.2	1881	21	AAV44506
41	113.5	4.2	494	22	ABBS5808
42	113.5	4.2	494	22	ABBS5808
43	112.5	4.1	410	22	ABG01159
44	111	4.1	1036	19	AAW53280
45	110.5	4.1	383	23	ABW94357

#### ALIGNMENTS

RESULT 1	AAV75753
ID	AAV75753 standard; Protein: 508 AA.
XX	AAV75753;
AC	AAV75753;
XX	AAV75753;
DT	21-MAR-2000 (first entry)
XX	AAV75753;
DE	Neisseria meningitidis ORF 987 protein sequence SEQ ID NO:2978.
XX	AAV75753;
KW	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW	antibacterial; gene therapy.
XX	AAV75753;
OS	Neisseria meningitidis.
XX	AAV75753;
PN	WO9957280-A2.
XX	AAV75753;
PD	14-MAY-1999.
XX	AAV75753;
PE	30-APR-1999; 99WO-US09346.
XX	AAV75753;
PR	01-MAY-1998; 98US-0083758.
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0098994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	AAV75753;
PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.

H. pylori ORF 05cp  
Staphylococcus aur  
Staphylococcus aur  
Staphylococcus epi  
Corynebacterium g1  
C. glutamicum prote  
C. glutamicum ATCC  
Lactococcus lactis  
S. epidermidis ope  
Propionibacterium  
Novel human diagno  
Novel human diagno  
Streptococcus poly  
Streptococcus poly  
Streptococcus poly  
Human phospholipase  
Chlamydia pneumonia  
S. cinamomeus prot  
Human phospholipase  
Human phospholipase  
Human phospholipase  
Chlamydia trachoma  
Marine phospholipase  
Moraxella catarrha  
Human phospholipase  
Human phospholipase  
Human phospholipase  
M. catarrhalis MCAL  
Herbivore acetyl  
Streptococcus pneu  
Drosophila melanog  
Novel human diagno  
Human phospholipase  
Chlamydia trachoma

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelein H, Venter JC;  
 XX  
 DR WPI: 2000-062150/05.  
 DR N-PSDB: AA254515.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 2: Page 1396-1397; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria bacteria* (e.g. meningitis and septicemia), to detect the  
 CC presence of *Neisseria bacteria*, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have also be used in gene therapy protocols. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX Sequence 508 AA:

Query Match 96.9%; Score 2629; DB 21; Length 508;  
 Best Local Similarity 100.0%; Pred. No. 2,76-241;  
 Matches 508: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 18 MTRSLISLILCLLSCSSWLPLEERTSRHFNTPSKVRLDNIQIRHTPTNGLSIDY 77  
 DB 1 MTRSLISLILCLLSCSSWLPLEERTSRHFNTPSKVRLDNIQIRHTPTNGLSIDY 60  
 OY 78 LNDPHEAFARAAALIESAHSLSLDQYIWRNDISGRLLFNLYVLAARGVRRLLDDN 137  
 DB 61 LNDPHEAFARAAALIESAHSLSLDQYIWRNDISGRLLFNLYVLAARGVRRLLDDN 120  
 OY 138 NTRGLDDLLALDSHPNIEVLFNFVLRKWRALGYLDPRLNRMNKSFTADNRATI 197  
 DB 121 NTRGLDDLLALDSHPNIEVLFNFVLRKWRALGYLDPRLNRMNKSFTADNRATI 180  
 OY 198 ICGRNIGDEYFVGDYFADIDILATGSGVGEVSHDPRYASAHNATIIISGNTG 257  
 DB 181 ICGRNIGDEYFVGDYFADIDILATGSGVGEVSHDPRYASAHNATIIISGNTG 240  
 OY 258 KGLQALGYNDETSRHALLYRETVEOSPLYOIKIOTGRIDMOSVQIRLISDDPAKGLDRDR 317  
 DB 241 KGLQALGYNDETSRHALLYRETVEOSPLYOIKIOTGRIDMOSVQIRLISDDPAKGLDRDR 300  
 OY 318 RKPPIAGRLQDALKOPERSVLYSPYFVPTKSGTDALALVODGIDVTVLNSLOATDVA 377  
 DB 301 RKPPIAGRLQDALKOPERSVLYSPYFVPTKSGTDALALVODGIDVTVLNSLOATDVA 360  
 OY 378 AVHSGYVYKRPRLKAGIKLYELOPNHNAPAKDKGLTSSVTSIHAFTFYDGKRIFIG 437  
 DB 361 AVHSGYVYKRPRLKAGIKLYELOPNHNAPAKDKGLTSSVTSIHAFTFYDGKRIFIG 420  
 OY 438 SFNLDPRSARLNTGCVYTESPKIAEOMERTLADTSPEYAYVTLDHRHRLQWHPATRK 497  
 DB 421 SFNLDPRSARLNTGCVYTESPKIAEOMERTLADTSPEYAYVTLDHRHRLQWHPATRK 480  
 OY 498 TYPNEPEAKLMKRIAKILSLPIESL 525  
 DB 481 TYPNEPEAKLMKRIAKILSLPIESL 508

RESULT 2  
 AA275752  
 ID AA275752 standard; Protein: 508 AA.

XX AA275752;  
 AC  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE *Neisseria meningitidis* ORF 987 protein sequence SEQ ID NO:2976.  
 XX  
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy.  
 XX  
 OS *Neisseria meningitidis*.

XX W09957280-A2.  
 XX 11-NOV-1999.  
 XX 30-APR-1999; 99WO-US09346.  
 XX 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.

PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelein H, Venter JC;  
 DR WPI: 2000-062150/05.  
 DR N-PSDB: AA254514.

PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics

PS Claim 2: Page 1395; 1453pp; English.

CC AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria bacteria* (e.g. meningitis and septicemia), to detect the  
 CC presence of *Neisseria bacteria*, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have also be used in gene therapy protocols. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX Sequence 508 AA:

Query Match 95.8%; Score 2599; DB 21; Length 508;  
 Best Local Similarity 98.8%; Pred. No. 1,9e-238;  
 Matches 502: Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 18 MTRSLISLILCLLSCSSWLPLEERTSRHFNTPSKVRLDNIQIRHTPTNGLSIDY 77  
 DB 1 MTRSLISLILCLLSCSSWLPLEERTSRHFNTPSKVRLDNIQIRHTPTNGLSIDY 60  
 OY 78 LNDPHEAFARAAALIESAHSLSLDQYIWRNDISGRLLFNLYVLAARGVRRLLDDN 137  
 DB 61 LNDPHEAFARAAALIESAHSLSLDQYIWRNDISGRLLFNLYVLAARGVRRLLDDN 120  
 OY 138 NTRGLDDLLALDSHPNIEVLFNFVLRKWRALGYLDPRLNRMNKSFTADNRATI 197



Db 121 NTRGLDILLALDSDHPNIEVLFNPFVLRKMRALGYLTDFPRLRRMHNKSFTADNRAT 180  
 QY 198 LGGNIDEXFKVGEDYFADLDILATGSGVGEVSHDFDXYMASHAHNATRIIRSGNIG 257  
 Db 181 LGGNIDEXFKVGEDYFADLDILATGSGVGEVSHDFDXYMASHAHNATRIIRSGNIG 240  
 QY 258 KGLQALGYNDTSRHALRYRETVESQPLVQKIQTGRIIDMOSVQVTLISDDPAKGLDRDR 317  
 Db 241 KGLQALGYNDTSRHALRYRETVESQPLVQKIQTGRIIDMOSVQVTLISDDPAKGLDRDR 300  
 QY 318 RKPPIAGRLQDALQKQPKSVYLVSPYFVPRKSGTDALAKLVODGIDVTVLTNSLQATDVA 377  
 Db 301 RKPPIAGRLQDALQKQPKSVYLVSPYFVPRKSGTDALAKLVODGIDVTVLTNSLQATDVA 360  
 QY 378 AVHSGYVYKRRPLKAGIKIYELQPNHNAVPAATKDKGLTGSSVYSLHAKFTIYDGRKIFIG 437  
 Db 361 AVHSGYVYKRRPLKAGIKIYELQPNHNAVPAATKDKGLTGSSVYSLHAKFTIYDGRKIFIG 420  
 QY 438 SFNIDPRSARLNTMGVIESPKIAEQMERTLADTSPRYAVRYTLDRHNLQWHDPAATRK 497  
 Db 421 SFNIDPRSARLNTMGVIESPKIAEQMERTLADTSPRYAVRYTLDRHNLQWHDPAATRK 480  
 QY 498 TYRNEPEAKLMKRIAATILSLPIESLL 525  
 Db 481 TYRNEPEAKLMKRIAATILSLPIESLL 508

## RESULT 3

AAV75751  
 ID AAV75751 standard; Protein; 507 AA.

AC AAV75751;

DT 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 987 protein sequence SEQ ID NO:2974.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;

KM antibacterial; gene therapy.

XX OS Neisseria gonorrhoeae.

XX PN MO9957280 =A2.

XX PD 11-NOV-1999.

XX PR 30-APR-1999; 99WC-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098962.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR ) CHIRON CORP.

XX PI (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masjanani V, Mora M;  
 XX PI Petersen J, Plaza M, Rappoldi R, Ratti G, Scalato E, Scarselli M;  
 XX PI Tettelin H, Venter JC;  
 XX DR WPI: 2000-062150/05.  
 XX DR N-PSDB; AA254513.  
 XX PT Novel Neisserial polypeptides predicted to be useful antigens for  
 XX PS vaccines and diagnostics  
 XX PS Claim 2; Page 1394; 1453pp; English.

CC AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254733 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC *Neisseria meningitidis* (e.g. meningitis and septicemia), to detect the  
 CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have also as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX SO Sequence 507 AA;

Query Match 95.08; Score 2577.5; DB 21; Length 507;  
 Best Local Similarity 98.28; Pred. No. 2,1e-236;  
 Matches 499; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 18 MKTRSLISLCLLCSCSWMPLPEERTESHNTSKPRVLDNLTQIRHTPTNGSLDIY 77

Db 1 MKTRSLISLCLLCSCSWMPLPEERTESHNTSKPRVLDNLTQIRHTPTNGSLDIY 60

QY 78 LINDPHEAFARAALIESAEHSIDLYIWRNDISGRLEFNLVYLAERGVRRLLDDN 137

Db 61 LINDPHEAFARAALIESAEHSIDLYIWRNDISGRLEFNLVYLAERGVRRLLDDN 120

QY 138 NTRGLDILLALDSDHPNIEVLFNPFVLRKMRALGYLTDFPRLRRMHNKSFTADNRAT 197

Db 121 NTRGLDILLALDSDHPNIEVLFNPFVLRKMRALGYLTDFPRLRRMHNKSFTADNRAT 179

QY 198 LGGNIDEXFKVGEDYFADLDILATGSGVGEVSHDFDXYMASHAHNATRIIRSGNIG 257

Db 180 LGGNIDEXFKVGEDYFADLDILATGSGVGEVSHDFDXYMASHAHNATRIIRSGNIG 229

QY 258 KGLQALGYNDTSRHALRYRETVESQPLVQKIQTGRIIDMOSVQVTLISDDPAKGLDRDR 317

Db 240 KGLQALGYNDTSRHALRYRETVESQPLVQKIQTGRIIDMOSVQVTLISDDPAKGLDRDR 299

QY 318 RKPPIAGRLQDALQKQPKSVYLVSPYFVPRKSGTDALAKLVODGIDVTVLTNSLQATDVA 377

Db 300 RKPPIAGRLQDALQKQPKSVYLVSPYFVPRKSGTDALAKLVODGIDVTVLTNSLQATDVA 359

QY 378 AVHSGYVYKRRPLKAGIKIYELQPNHNAVPAATKDKGLTGSSVYSLHAKFTIYDGRKIFIG 437

Db 360 AVHSGYVYKRRPLKAGIKIYELQPNHNAVPAATKDKGLTGSSVYSLHAKFTIYDGRKIFIG 419

QY 438 SFNIDPRSARLNTMGVIESPKIAEQMERTLADTSPRYAVRYTLDRHNLQWHDPAATRK 497

Db 420 SFNIDPRSARLNTMGVIESPKIAEQMERTLADTSPRYAVRYTLDRHNLQWHDPAATRK 479

QY 498 TYRNEPEAKLMKRIAATILSLPIESLL 525

Db 480 TYRNEPEAKLMKRIAATILSLPIESLL 507

## RESULT 4

AAW55452  
 ID AAW55452 standard; Protein; 502 AA.

AC AAW55452;

DT 24-JUN-1998 (first entry)

DE H. pylori ORF 02ae11612\_33203250\_c1\_51 secreted protein.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;

KW identification; binding compound; bacteria; life cycle; activator;

KM inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

XX OS Helicobacter pylori.



CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.

XX Sequence 502 AA;

Query Match 23.3%; Score 632; DB 18; Length 502;  
 Best local similarity 32.9%; Pred. No. 4,6e-51;  
 Matches 163; Conservative 92; Mismatches 184; Indels 56; Gaps 17;

QY 64 IRTPTNGISDIY-----LINDPEAFARALIESAHSLSLDQYIYWRN 109  
 DB 30 ISDPTTTTIGSLYAKNKKENPKHSAAILLEDGFDAHLHVGGLIRMSQKSIDQYIYKN 89  
 QY 110 DISGRLLFNLYLAERGVNRLLDDNTRGID---DLLALDSRNIEVLFNFVYL 165  
 DB 90 DISSQYIAKELLANNGVNRILLDDN--GLDSFSDIML-LNPKNIEVIFNPFYI 145  
 QY 166 RKRRALGY--LTDPRLRNRHMKSFADNRATILGRNIGDEYFVGEFVADLIDL 222  
 DB 146 RN-KGLRYFEMLADYERIKRHMKNKLFYDNFAVIIGRNIGDNYFDNDJTNFLDIDL 204  
 QY 223 ATGSVYGEVSHDPRYASHAHNATRIIRS-----GNICKGLQALGYNDTSRHAL 274  
 DB 205 FFGGVASKAKESFENYWRFRHSIPVS-LRTHKRLKNVYEIAKLEKIPISADANEFE 263  
 QY 275 LRHREVEOSPLYO-KIOTGRIMOSVQTRLLSDPAKGRDRRRPRPIAGRLQDADKOP 333  
 DB 264 KYNNDIEROKYOPIYYG-----NAIFLADLPK-IDTPLYS-PIKIAFEKAKNA 314  
 QY 334 EKSIVYLVSPYFVPTKSGTDALAKLVODGVTLTNSLQATDVAAVHSGVYKRPPLKA 393  
 DB 315 KDSVFIASSYFIFGKIMKIFKNQISGIELNLTNSLSTDAIVYGAERIRNKLVRM 374  
 QY 394 GILYELQPNHAYPATKDKGLTSSVTSIAKFFIYDGKFIIGSENLDPKRLNTEMG 423  
 DB 375 GANVYELRNDFFNRQIKGR--FSTKHSIHKGTIVFDALTLTSGFNIDPRSAVINTESA 431  
 QY 454 IVIESPRIAOMERTLADTSPEYAYRYTLDRHNRLOMHDPATRKY----PNEPEAKLMK 509  
 DB 432 VLDNDSFARKVRSLKDAH-QQSWHLVLYRH-RVIME--ATEBGILLIHEKNSPDYSFL 487  
 QY 510 RIAAKILSLPIESL 524  
 DB 488 RLKEMSKVLPEREL 502

RESULT 6  
 AAM98378  
 ID AAM98378 standard; Protein; 428 AA.  
 XX  
 AC AAM98378;  
 XX  
 DT 31-MAR-1999 (first entry)  
 XX  
 DE H. pylori GHP0 1375 protein.  
 XX  
 KM GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;  
 KM peptic ulcer disease.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO9843478-A1.  
 XX  
 PD 08-OCT-1998.  
 XX  
 PF 01-APR-1998; 98WO-US06371.  
 XX  
 PR 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 XX

PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;

DR WPI; 1998-542293/46.  
 DR N-PSDB; AAX14097.

PT New isolated Helicobacter polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of Helicobacter  
 PT infections and gastrointestinal diseases

Claim 8; Page 581-583; 2054pp; English.

This sequence represents a Helicobacter pylori GHP0 protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC Helicobacter infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.

SQ Sequence 428 AA;

Query Match 21.8%; Score 591; DB 19; Length 428;  
 Best local similarity 32.2%; Pred. No. 2,8e-47;  
 Matches 150; Conservative 87; Mismatches 153; Indels 76; Gaps 17;

QY 96 AEHSLSLDQYIYWRNLDISGRLLFNLYLAERGVNRLLDDNTRGID---DLLALDS 151  
 DB 2 SOKSIDMQYIYWRNLDISSQYIAKELLANNGVNRILLDDN--GLDSFSDIML-LNF 57  
 QY 152 HPNIEVLFNPFYLRKRRALGY--LTDPRLRNRHMKSFADNRATILGRNIGDEYF 208  
 DB 58 HKNIEVIFNPFYIYRN-KGLRYFEMLADYERIKRHMKNKLFYDNFAVIIGRNIGDNYF 116  
 QY 209 KVEDIVFADLIDLATGSSVYGEVSHDPRYASHAHNATRIIRSGNIGKGLALGYND 268  
 DB 117 DNDJTNFLDIDLFCGVASKAKESPERYWRFRHSIPVS-LRTH----- 160  
 QY 269 TSKHALLRYEYEOSPLYOKIOTGRIMOSVQTRL-----ISD 307  
 DB 161 ---HKRLK-NNAKIEIAKLEKIPISADKNOFEKVDIDRQKYOPIYYGNALFLAD 216  
 QY 308 DPAKGLDRDRKPPDIAGRLDALDKOPKESVYLVSPYFPTKSGTDALAKLVODGVTVL 367  
 DB 217 SPKK-IDTPLYS-PIKIAFEKALKNAKDSVFIASSYIFPKKMMKIFKNQISGIELNLT 274  
 QY 368 TNSLQATDVAAVHSGVYKRPPLKAGIKLYELOPNHAYPATKDKGLTSSVTSIAKTF 427  
 DB 275 TNSLSTDAIVYGAERIRNQLVRMGANYEIRNDEFNRQIKGR--FSTKHSIHGKTI 331  
 QY 428 IVDGKRIFGSEFNIDPRSAALNTEMGVIESPKIAEQMERTLADTSPEYAYRYTLDRHNR 487  
 DB 332 VFDDNLTLLGSEFNIDPRSAVINTESAVALFDNPSFAKVRSLKDAH-QQSWHLVLYRH-R 389  
 QY 488 LOW-----HDPATRKTYPNPEEAKLMKRIAATILSLPIESL 524  
 DB 390 VIMEAVEGILIH-----KT---SPDTSFRLRLIKEMSKVLPEREL 428

RESULT 7  
 ABB49813  
 ID ABB49813 standard; Protein; 504 AA.  
 XX  
 AC ABB49813;  
 XX  
 DT 05-FEB-2002 (first entry)  
 XX  
 DE Listeria monocytogenes protein #2517.  
 XX  
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
 KM vitamin B12; bacterial infection; disease.



[illegible]

PI Buchrieser C, Frangoul L, Couve E, Rusniok C, Eshti H, Dehoux P,  
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P,  
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA,  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amond A,  
PI Chakraborty T, Domm E, Hain T, Berche P, Chabibit A, Durant L,  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
PI Madueno E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J,  
PI Rose M, Voss H;  
DR WPI, 2002-010914/01.  
XX  
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and  
PI related polypeptides  
XX  
XX  
PS Claim 6; SEQ ID NO 2698; 192pp; French.  
XX  
XX The present invention relates to the genome sequence of *Listeria*  
CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in *L.*  
CC *monocytogenes* and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of *L. monocytogenes* and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccine compositions for the treatment or prevention of infections by *L.*  
CC *monocytogenes* and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
CC  
XX  
XX  
SQ Sequence 482 AA;  
Query Match 10.6%; Score 288.5; DB 23; Length 482;  
Best Local Similarity 24.3%; Pred. No. 2.3e-18;  
Matches 112; Conservative 61; Mismatches 172; Indels 115; Gaps 16;  
OY 74 SDIYLLNDPHEAFARAAALIESAHSLSLDQYIWRNDISGLLENLYLAERGVAVRL 133  
DB 119 NEVELFVGDHEKFDALIDIEKAKDHILYIFHSDEGNRLMYLEKKAEGLVWKT 178  
OY 134 LDDNNTRGLDILLALDHPNIEVLEPNFVLRKWRALCYLTDFEPLNRMRNKSFTADN 193  
DB 179 YDAMGSRRTKKSFFRPFQKNGGLVRPFPSKLP-----LINF-RLNVRNHRKLAIIDG 230  
OY 194 RATILIGRNIQDEYFKVGEDYFADLLILATGVSVEGSHDPRVWASHANAFIITS 253  
DB 231 DVGYYGGEFYGDEY--LGASKRFG-----WRD-----THLRVH 262  
OY 254 GNIGKGLA---LGYNDERSHAL---LRYRETV---EGLPYOKIOTG-RIDMOSVOTRL 304  
DB 263 GAAVYAMQTRFLIMDNNSASTIKIDYKAKYFTFPHGKGTSMQIYSSGSDSEMOQIKNTY 322  
OY 305 ISDDPAKGLDRDRRPPRAGLQDALIKOPKESVYLVSPYFVTKSGDTALVODGIDV 364  
DB 323 IK-----MINKAKTIYLOSPPFIDASLEAIKIALISGVTV 360  
OY 365 TYLTNSLQATDYAAVHSGVYKRLKAGIKLYELQPHNAVPAKTKDKLTGSSVTSLHA 424  
DB 361 RMIWN--KPDHAFYRATTNAGELMEGAKIF-----YDNGF-----IHA 401.  
OY 425 KTFIWDGRIFGSEFNLDPRSARLNTMGVLESPIKIAEMERTLADTSPETAYAVTIDR 484  
DB 402 KTLVVDGELASVGTANMDPSRFLNFEVNAFYERKOMVKEDADLED----- 449  
OY 485 HNRLOMHPDPRKTYPNPE---AKLW---KRIAKIIS 517  
DB 450 -----ILKSQYLPPELYAKRSIMWIKKEAVSRLLS 479

RESULT 11  
ID AAM55249  
AAM55249 standard; Protein; 206 AA.  
XX  
XX AAM55249;  
AC  
XX  
XX 26-JUN-1998 (first entry)  
DE H. pylori ORF 05cp1191lorf35 protein.  
XX  
XX  
KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
KW identification; binding compound; bacteria; life cycle; activator;  
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
KW bacterium.  
XX  
XX  
OS Helicobacter pylori.  
XX  
XX W09737044-A1.  
XX  
XX 09-OCT-1997.  
XX  
XX 27-MAR-1997; 97WO-US05223.  
XX  
XX 06-DEC-1996; 96US-0761318.  
XX 29-MAR-1996; 96US-0625811.  
XX 02-APR-1996; 96US-0758731.  
XX 25-OCT-1996; 96US-0736905.  
XX 28-OCT-1996; 96US-0738859.  
XX  
XX (ASTR ) ASTRA AB.  
XX  
XX Alm RA, Smith D;  
PI WPI, 1997-503122/46.  
DR N-PSDB; AAV24658.  
XX  
XX Helicobacter pylori nucleic acid sequences and encoded  
PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
PT infection and for diagnosis of H. pylori infection  
XX  
XX  
PS Claim 14; Page 492; 1145pp; English.  
XX  
XX This sequence is a Helicobacter pylori protein of unspecified  
CC function.  
CC The protein may be used in a vaccine to prevent or treat H. pylori  
CC infection or to identify H. pylori polypeptide binding compounds,  
CC useful as potential H. pylori life cycle activators or inhibitors.  
CC The DNA and probes derived from it may be used for the  
CC identification of H. pylori in a sample, and the diagnosis of  
CC H. pylori infection. Nucleic acid sequences complementary to the  
CC DNA act as antisense sequences, and can be used to prevent the  
CC translation of H. pylori mRNA. Antibodies against the protein can  
CC be used in immunoassays to evaluate the abundance and distribution  
CC of H. pylori-specific antigens. The genomic sequence of H. pylori  
CC (ATCC 55679) was determined from overlapping contigs generated by  
CC mechanically shearing the bacterial DNA. The sequences were  
CC analysed for ORF of at least 180 nucleotides, and the predicted  
CC coding regions defined by computer evaluation. To identify likely  
CC H. pylori antigens for vaccine development, the amino acid  
CC sequences predicted from various ORF were analysed for significant  
CC homology to other known or exported membrane proteins. Having  
CC identified and determined the sequences of interest, particular  
CC regions can be isolated from H. pylori by PCR amplification for  
CC recombinant polypeptide production, e.g. in E. coli hosts.  
XX  
XX  
SQ Sequence 206 AA;  
Query Match 10.3%; Score 279.5; DB 18; Length 206;  
Best Local Similarity 31.7%; Pred. No. 4.2e-18;  
Matches 66; Conservative 45; Mismatches 86; Indels 11; Gaps 5;

Query	Local Match	Score	DB	Length
321	PIAGLQALAKQPEKSVYLVSYPFPTSGTDLAKLVODGIDVPLTSLQADVAAYH	380		
321	PIAGLQALAKQPEKSVYLVSYPFPTSGTDLAKLVODGIDVPLTSLQADVAAYH	380		
6	PIKIAFEALAKNAKSVFLASTYFIPGKKIMKIFKNQISKGLIELTSLSTDAIVY	65		
381	SGVYKRRPLKAGIKLVELQPNHVAVPATKDKGLTGSVYSLHAKTFIVDGKRIFGSPN	440		
66	GAMERYRKLKLVGMGANVYEIRFNDPFNRQIKGR--FSTKHSIHGKTIIVDDALLTSGFN	122		
441	LDPRSARLNTGEGVIVIESPKIAEQMERTLADISPEAVAVTLDRNRLQMHDPATKTY	499		
123	IDPRSAVYNTGESAVLEFNDNSFAKRYRLSLKHA-QDSWMLVLYRH-RVIWE--ATEEGIL	178		
500	--PNEPEAKLMKRIAATKLSLPIESL	524		
179	IHEKNSPDTSEFLRLIKEMSKVLPREL	206		
RESULT 12				
ABB32481				
ID	ABB32481 standard; Protein: 493 AA.			
AC	ABB32481;			
DT	23-JAN-2002 (first entry)			
DE	Staphylococcus aureus polypeptide SEQ ID NO 22.			
DE	Staphylococcus aureus; antimicrobial; growth rate; infection; vaccine;			
IM	immunity.			
XX	Staphylococcus aureus.			
XX	WO200177365-A2.			
PD	18-OCT-2001.			
PF	06-APR-2001; 2001WO-US11177.			
XX	06-APR-2000; 2000US-266327P.			
PA	(PHAA ) PHARMACIA & UPJOHN CO.			
PI	Tomlch CC;			
DR	WPI: 2002-010920/01.			
XX	N-PSDB: ABA51633.			
PT	Identifying antibacterial agents which bind to polypeptide encoded by			
PT	essential coding sequences of Staphylococcus aureus, and which reduce			
PT	growth rate of bacteria, by determining binding of agent to the			
XX	polypeptide			
PS	Example 1; Fig 1; 67pp; English.			
CC	The invention relates to identifying antimicrobial agents that target the			
CC	proteins (ABB32471-ABB32494) encoded by genes (ABA51633-ABA51634 and			
CC	ABA51659-ABA51670) essential for the survival of Staphylococcus bacteria,			
CC	especially Staphylococcus aureus, including antimicrobial agents that			
CC	interfere with the expression of essential coding sequence products and			
CC	antimicrobial agents that interfere with the function of essential coding			
CC	sequence products. The genes and proteins are useful for identifying			
CC	agents that decreases the growth rate of a microbe preferably S. aureus.			
CC	The agents identified by the above method are bactericidal and will be			
CC	useful for treating a subject infected with S. aureus or at risk of being			
CC	infected by S. aureus and may be used in vaccines for producing			
CC	immunological response to S. aureus resulting in immunity. The present			
CC	sequence is that of a protein encoded by an essential S. aureus			
CC	polynucleotide coding sequence.			
XX				
SO	Sequence 493 AA;			
Query Match	10.1%; Score 275; DB 23; Length 493;			
Best Local Similarity	21.5%; Pred.No. 4.5e-17;			

Matches	100:	Conservative	81:	Mismatches	149:	Indels	136:	Gaps	16
Qy	74	SDIY--LINDPHEAFARAAALIESAEHSIDLOYYIWRNDISGRLEFNLYLAAERGVRYR	131						
Db	139	NDLYQLVKLD-----	187						
Qy	132	LLLDNNFRGGLDLLLLALDSHPNIEVRLFPVLRKWRALG-----	181						
Db	188	ILYDVGSK-----	229						
Qy	182	RRMHNKSEFADNRATILGGRNIGDEXEFKGEDTVEADLDIILATGSVYGEVSHDFDRYMAS	241						
Db	230	NRNHRKIIVDQGLGVYGFNGIDELGLGKLGWMDTLRIQGDVADVALQLRFILDMNS	289						
Qy	242	HSANHATRI-----	294						
Db	290	-QAHNPOFEYDVVKYEPKKNGPLG-----	323						
Qy	295	IDMOSVQ---TRLIEDDPAKGLDRDRRKPPIAGRLQDALOPEKSVLYLSPYFVPTKSGT	351						
Db	324	SDMHQIEGYTFMI-----	358						
Qy	332	DALAKLVODGDIDYVLTNSLQATDYAAVHSGVYKTRKPLLKAGIKLYELQPNHAPATKD	411						
Db	359	NAIKAAKSGVDVHLMIPC--KPDHPLVYWATFSNASDLLSGVKIY-----	406						
Qy	412	KGLTGSVYSLCAKTFEIVDGKRIFGFNLDRSALNTNEMGVIESPKIAEOMERTLAD	471						
Db	407	NGF-----	453						
Qy	472	TSPEAYAYRVTLDRHNRLQWMDPATRKTYYPNEBEAKIMKRIAKILS	517						
Db	454	--FAVEHDITRSKQL-----	490						

RESULT 13

ABB32493

ID

ABB32493 standard; Protein; 502 AA.

ABB32493;

23-JAN-2002 (first entry)

DE

Staphylococcus aureus polypeptide SEQ ID NO 90.

XX

Staphylococcus aureus; antimicrobial; growth rate; infection; vaccine; immunity.

OS

Staphylococcus aureus.

XX

WO200177365-A2.

XX

18-OCT-2001.

XX

06-APR-2001; 2001WO-US11177.

XX

06-APR-2000; 2000US--266327P.

XX

(PHAA ) PHARMACIA & UPJOHN CO.

XX

Tomlich CC;

XX

WPI; 2002-010920/01.

XX

N-PSDB; ABA51669.

XX

Identifying antibacterial agents which bind to polypeptide encoded by essential coding sequences of Staphylococcus aureus, and which reduce growth rate of bacteria, by determining binding of agent to the polypeptide

XX

Example 2; Fig 2; 67pp; English.

XX

The invention relates to identifying antimicrobial agents that target the

CC proteins (ABR32471-ABR32494) encoded by genes (ABA51623-ABA51634 and  
 CC ABA51659-ABA51670) essential for the survival of *Staphylococcus bacteria*,  
 CC especially *Staphylococcus aureus*, including antimicrobial agents that  
 CC interfere with the expression of essential coding sequence products and  
 CC antimicrobial agents that interfere with the function of essential coding  
 CC sequence products. The genes and proteins are useful for identifying  
 CC agents that decrease the growth rate of a microbe preferably *S. aureus*.  
 CC The agents identified by the above method are bactericidal and will be  
 CC useful for treating a subject infected with *S. aureus* or at risk of being  
 CC infected by *S. aureus* and may be used in vaccines for producing  
 CC immunological response to *S. aureus* resulting in immunity. The present  
 CC sequence is that of a protein encoded by an essential *S. aureus*  
 CC polynucleotide coding sequence for expression in *E. coli*.

SO Sequence 502 AA:

Query Match 10.1%; Score 275; DB 23; Length 502;  
 Best Local Similarity 21.5%; Pred. No. 4.7e-17;  
 Matches 100; Conservative 81; Mismatches 149; Indels 136; Gaps 16;

QY 74 SDIY--LNDPHEAFARALIESAHSIDLOYIRNDISGRILFNLYLAERGVRRLD 131  
 DB 140 NDLVDVLDK-----IKMAEYIHLEYTFALDGLKRIILMALEKLEKQGLEVK 188  
 DB 132 LLLDNNTEGLDLDLALDHSNIEVRLFPYLRKRALG-----YLTDP--RLN 181  
 DB 189 ILYDDVGSK-----NVKMANDFH-----KSLGGEVEAFPSKPLPLFRNM 230  
 QY 182 RMHMKSFADNRATILGSRNIGDEYKVEDTVEADLILATGSVGVSHDFRYMAS 241  
 DB 231 NNNHRIIYIDQLGVYGGFNIGDEYLGKLGWRDTHLRIGDPAVDALQLRFLDMNS 290  
 QY 242 HSAHNATRII--RSGNIGKGLQALGYNDTSRHALLRYREVEOSPXYOKOTR 294  
 DB 291 -QAHNPOFEYDVKKPKKNGPLG-----NSPQIYASGPA 324  
 QY 295 IDMOVQ--TRLISDDPAKGLDRDRKRPPIAGRLQDALKEPSYLVSPFVPTKSGT 351  
 DB 325 SDMHQIEGYTKMI-----MSAKKSYLQSPFIFDINSYI 359  
 QY 352 DALALVDGIDIVYLTNSLOATDYAAVHSGYVKRPLKAGIKLYELQPHNAVPAIKD 411  
 DB 360 NAIKIAASGVVHLMIPC--KPDHPLYWATFSSNASDLSGVSXIV-----TYE 407  
 QY 412 KGLTSSVYSLAKTEIYVDKRIFFGSFNLDRSARLNTMGVIESPKIAQOMERTIAD 471  
 DB 408 NCF-----HSKMLDDEIVSGTAMDRSEFELNFEVNAFYDEMIAKDL----- 454  
 QY 472 TSEPAAYRVLDRHNRLOWHDPATRTKTYPNPEAKLMKRIAKIIS 517  
 DB 455 ---RVAYEHDTLTKSKQL-----TKESYANRPLSVKFKSLAKLVS 491

RESULT 14

ABP39331  
 ID ABR39331 standard; Protein: 494 AA.

ABP39331;

24-JUL-2002 (first entry)

Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4176.

Staphylococcus epidermidis: open reading frame; ORF; bacterial infection;

antibacterial; gene therapy.

Staphylococcus epidermidis.

US6380370-B1.

30-APR-2002.  
 13-AUG-1998; 98US-0134001.

XX 14-AUG-1997; 97US-053779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 XX WPI: 2002-381255/41.  
 DR N-PSDB: ABR91876.  
 XX

PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PS Disclosure; SEQ ID 4176; 267pp; English.

CC ABR90538 to ABR93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABR3124 to ABR37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life  
 CC cycle or inhibit *S. epidermidis* infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.

SO Sequence 494 AA;

Query Match 9.7%; Score 264.5; DB 23; Length 494;  
 Best Local Similarity 22.3%; Pred. No. 4.6e-16;  
 Matches 101; Conservative 71; Mismatches 177; Indels 103; Gaps 11;

QY 76 IYLDNPEAFARALIESAHSIDLOYIRNDISGRILFNLYLAERGVRRLD 135  
 DB 133 IDLFTDGHKLYEKVLEDIYNAQDIHLEYTFELDGLKRIILDALETKKEGLEKLLXD 192  
 QY 136 DNNTEGLDLDLALDHSNIEVRLFPYLRKRALGYITD--FP-----RLNRMH 185  
 DB 193 DVGSK-----KVRLSK--FKHFRALGGEVEAFPSKVPYLINFRNNNRH 234  
 QY 186 NKSFTADNRATILGSRNIGDEYKVEDTVEADLILATGSVGVSHDFRYMASHAH 245  
 DB 235 KRIITIDQIGVGFNWDGLGKLGWRDTHLRIGDPAVDALQLRFLDMNSQS-- 292  
 QY 246 NATRIIRSGNIGKGLQALGYNDTSRHALLRYREVEOSPXYOKOTGRIDMOVSOTRLI 305  
 DB 293 -----HRPQFKDQKFPKNG--DKGNAIIDA 319  
 QY 306 SDDPAKGLDRDRKRPPIAGRLQDALKEPSYLVSPFVPTKSGTDALAKIVDGDIVT 365  
 DB 320 SSGPAFDLHQ-----IEGYTKMISAKKSIYLOSPFIPDOSYINALKMAANSVGEVN 373  
 QY 366 VLTNSLOATDYAAVHSGYVKRPLKAGIKLYELQPHNAVPAIKDKGLTSSVYSLAK 425  
 DB 374 LMIPC--KPDHPLYWATFSSNASDLSGVSXIVYQNF-----HSK 414  
 QY 426 TFIYDGRIFGSFNLDRSARLNTMGVIESPKIAQOMERTIADTSEPAAYRVLDRH 485  
 DB 415 ILMIDETISSIGSAMDRSEFELNFEVNAFYDEIDIAKOLQ-----AFEDIDQS 465  
 QY 486 NROLWHDPATRTKTYPNPEAKLMKRIAKIIS 517  
 DB 466 KLL-----TKEVYDKRPLSTIKFKGLAKLIS 491

RESULT 15

AAB76548  
 ID AAB76548 standard; Protein: 422 AA.

AAB76548;



DT 11-APR-2001 (first entry)  
 XX  
 DE Corynebacterium glutamicum MCT protein SPQ ID NO:78.  
 XX  
 XX Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;  
 KW membrane construction and membrane transport protein; petroleum spill;  
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
 KW identification; microorganism; fine chemical production; transformation;  
 KW genome mapping; genetic engineering.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100805-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00926.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 08-JUL-1999; 99DE-1031454.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031563.  
 PR 09-JUL-1999; 99DE-1033122.  
 PR 09-JUL-1999; 99DE-1033124.  
 PR 09-JUL-1999; 99DE-1033125.  
 PR 09-JUL-1999; 99DE-1033128.  
 PR 09-JUL-1999; 99DE-1033180.  
 PR 09-JUL-1999; 99DE-1033182.  
 PR 09-JUL-1999; 99DE-1033190.  
 PR 09-JUL-1999; 99DE-1033191.  
 PR 09-JUL-1999; 99DE-1033209.  
 PR 09-JUL-1999; 99DE-1033212.  
 PR 09-JUL-1999; 99DE-1033227.  
 PR 09-JUL-1999; 99DE-1033228.  
 PR 09-JUL-1999; 99DE-1033229.  
 PR 09-JUL-1999; 99DE-1033230.  
 PR 14-JUL-1999; 99DE-1033297.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040830.  
 PR 27-AUG-1999; 99DE-1040831.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 27-AUG-1999; 99DE-1040833.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 03-SEP-1999; 99DE-1041395.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042078.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 XX  
 PA (BADI ) BASF AG.  
 XX  
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;  
 XX  
 DR MPI: 2001-071486/08.  
 DR N-PSDB; AAF67781.  
 XX  
 XX Corynebacterium glutamicum nucleic acids encoding membrane construction  
 PT and membrane transport proteins or their portions, useful for typing or  
 PT identifying C. glutamicum or related bacteria, and as markers for  
 PT transformation -  
 XX  
 PS Claim 20: Page 294-296; 1119pp: English.  
 XX  
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
 CC construction and membrane transport (MCT) proteins given in AAB76510 to  
 CC AAB76847. The MCT nucleic acids and proteins are useful in the  
 CC identification of microorganisms which can be used to produce fine  
 CC chemicals, for modulating fine chemical production in C. glutamicum or

CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
 CC identification of C. glutamicum or related bacteria, as reference points  
 CC for mapping C. glutamicum genome, and as markers for transformation.  
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
 CC example from the present invention.  
 XX  
 XX  
 SQ Sequence 422 AA:  
 Query Match 9.0%; Score 243.5; DB 22; Length 422;  
 Best Local Similarity 22.9%; Pred. No.3.5e-14;  
 Matches 105; Conservative 61; Mismatches 179; Indels 113; Gaps 17;  
 QY 70 TNGLSIDYLLNDPEAFAPARALIESAHSIDQYIWRNDISGRLLFNLVYLAECGR 129  
 DB TCGNNGFY--SDYRESLKRMTAIDEAEYIYEIYIMADYSTQPFALERAHNNGVK 95  
 QY 130 VRLLD-----DNMTRLDLLALDSHPNIEVRNFNPFVLRKRKRALGYL 174  
 DB VRLFDHVGSKYRGYHRLKELRMGFAYLM-----LPLQPR----- 135  
 QY 175 TDPPRLNRMRHNSFTADNRATILGGRN-IGDEX-----FKVGDYFADLDIATGSV 228  
 DB RFRFRDRLNRHKKMLIIDGHTAFMGSONLIAPSLQKKNIKIGRE--WKDLMEITGPV 193  
 QY 229 GEVSHDFDRYWASHSAHNAIRIIRSGNIGKQALGYNDFTSRHALLRYRETYEQSPLYQ 288  
 DB SSMEMIFAGDWYVESN-----EALDIRDAEHAHGYIGNTQKDSATVLYQ 237  
 QY 289 KIOTGRIDWQSVQTRLISDPAKGLDRDRKRPPIAGLQDALQPEKSVLYVSPYPTK 348  
 DB LIPSG-----PGYTTEPNLR-----MENSIVHAKRRLILCSPYFIPDE 276  
 QY 349 SGTDALAKLVQDGDIDYVLTNSLOATDVAHVHGYVRRPLKAGIKLEYLPNHA VPA 408  
 DB SLEAVTSACRYGVYELFVS--EQADQFAIDHROSSYVALLAEAGKITQF--PKPV-- 331  
 QY 409 TKDKGLTSSVTSLHAKFTIVD----GKRIF--IGSFNLDPRSARLNTMGVIESPKI 461  
 DB LHTKYMIAIDPDPTGNALGLVIGSSNDIRSFGLNYEISLMIAGNL 378  
 QY 462 AEQMERTLADTSPFYARVTLDRKHNRLQWMDPATRKTY 499  
 DB 379 IHEL-NALTDRTYVTSFKLTLDKWNQRSW----RRRY 410  
 RESULT 16  
 AAG92760  
 ID AAG92760 standard; Protein; 500 AA.  
 XX  
 AC AAG92760;  
 XX  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 XX C glutamicum protein fragment SEQ ID NO: 6514.  
 DE  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EPI108790-A2.  
 PN  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 PA (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI: 2001-376931/40.

DR N-PSDB: AAH67979.

PT Novel polynucleotides derived from *Corynebacterium* bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 17: SEQ ID NO: 6514; 246bp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein  
CC sequences from the *Corynebacterium* *Corynebacterium* glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of *Corynebacterium* bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from *Corynebacterium* bacterium, and identifying a homologue of a gene derived  
CC from *Corynebacterium* bacterium. *Corynebacterium* bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

SO Sequence 500 AA:

Query Match 9.0%; Score 243.5; DB 22; Length 500;  
Best Local Similarity 22.9%; Pred. No. 4.7e-14;

Matches 105; Conservative 61; Mismatches 179; Indels 113; Gaps 17;

QY 70 TNGLSIDYLLNDPHEFAARALIESAHSIDLOYIWNDSIGRLFNLYIAERGV 129  
DB 116 TGGNNGFY--SDYRESLAKMTAIDEAEYIYELIMADSYTQPFALERAHNRGV 173  
QY 130 VRLDLD-----DNTRGDLDDLALDSHPNIEVRLNPNVLRKRAALGYL 174  
DB 174 VRLFDHVGSMKYPGYHRLKKELRMGFAYL-----LPLQPMR----- 213  
QY 175 TDFPRRLNRHMKSTADNRATILGRN-IGDEY-----FKYGEDTVFADLILATGSV 228  
DB 214 RFRFRDLNRHMKMLIDHTAFMGSQNLIPASYLOKNIKIGRE--WKDLWELTGPV 271  
QY 229 GEVSHDFRMYWASHAHNATRIIRSGNIGKGLAOLGYNDTSRHALLRRETVESPLQ 288  
DB 272 SSMEMIFADWYVESN-----EALDIRHAERHGTIGTQKDSATNLVQ 315  
QY 289 KIOTGRIDMOSVOTRLISDDPAKGLDRDRKRPPIAGRIODALKOPKESVYLVSPYFVPTK 348  
DB 316 LIPSG-----PGYTTEPNLR-----MENSIVHNAKERILICSPYFIPDE 354  
QY 349 SGTDLAKLYODGIDVYLTNSLOATDVAANVSGYKRPPLKAGITLYELOPHNAVPA 408  
DB 355 SLEAVTACYRGVTVLEFVS--EQADQFAIDHAOSSYQALLGAVKIYOF--PKPDV-- 409  
QY 409 TKDKGLTGSSVYSLAKTFIVD-----GKRIF--IGSFNLDPRSARLNTKMGVIESPKI 461  
DB 410 -----LHFKYMIADPPDPTGTGNEALGVLGSSNLDIRSGFLNTEISLMAKGNL 456  
QY 462 AEOMERTIADTSPEYAYRVTLDRHNRLQWHDPRATKTY 499  
DB 457 IHEL-NALTDRTYVSFKLTLDKMNQORSW-----RRRY 488

RESULT 17  
AAG80237  
ID AAG80237 standard; Protein; 500 AA.

XX AAG80237;

XX 25-JAN-2002 (first entry)

DE C. glutamicum ATCC13032 cIs protein.

KM Cardiolipin synthase; cIs gene; *Corynebacterium* L-glutamate;  
KM amino acid production; human medicine; animal nutrition; food additive.

XX *Corynebacterium* glutamicum.

PN DE10021826-A1.

XX 08-NOV-2001.

XX 04-MAY-2000: 2000DE-1021826.

XX 04-MAY-2000: 2000DE-1021826.

PA (DEGS ) DEGUSSA AG.

XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.

PI Nampoothiri M, Moeckel B, Pfeifferle W, Eggeling L, Sahm H;

DR WPI: 2002-027316/04.

DR N-PSDB: AAI68810.

PT New mutant *Corynebacterium* bacterium, useful for production of amino acids,  
PT especially glutamate, has increased activity of cardiolipin synthase  
XX  
XX  
XX  
PS Claim 6: Page 12-13; 20pp; German.

CC This invention describes a novel genetically modified *Corynebacterium*  
CC bacterium (A) in which activity of the cardiolipin synthase (cIs) gene,  
CC encoding cardiolipin synthase, is increased. (A) are used for  
CC fermentative production of amino acids, especially L-glutamate, useful in  
CC human medicine, animal nutrition, pharmaceuticals and as food additives.  
CC Nucleic acid, or its fragments, derived from the cIs gene are used (1) as  
CC primers for polymerase chain reaction of cIs genes or (11) as probes for  
CC isolating cDNA and genes that are highly homologous with cIs.  
CC Overexpression of the cIs gene results in increased production of amino  
CC acids. This sequence represents the *Corynebacterium* glutamicum ATCC13032  
CC cIs gene described in the method of the invention.

SO Sequence 500 AA:

Query Match 9.0%; Score 243.5; DB 23; Length 500;  
Best Local Similarity 22.9%; Pred. No. 4.7e-14;

Matches 105; Conservative 61; Mismatches 179; Indels 113; Gaps 17;

QY 70 TNGLSIDYLLNDPHEFAARALIESAHSIDLOYIWNDSIGRLFNLYIAERGV 129  
DB 116 TGGNNGFY--SDYRESLAKMTAIDEAEYIYELIMADSYTQPFALERAHNRGV 173  
QY 130 VRLDLD-----DNTRGDLDDLALDSHPNIEVRLNPNVLRKRAALGYL 174  
DB 174 VRLFDHVGSMKYPGYHRLKKELRMGFAYL-----LPLQPMR----- 213  
QY 175 TDFPRRLNRHMKSTADNRATILGRN-IGDEY-----FKYGEDTVFADLILATGSV 228  
DB 214 RFRFRDLNRHMKMLIDHTAFMGSQNLIPASYLOKNIKIGRE--WKDLWELTGPV 271  
QY 229 GEVSHDFRMYWASHAHNATRIIRSGNIGKGLAOLGYNDTSRHALLRRETVESPLQ 288  
DB 272 SSMEMIFADWYVESN-----EALDIRHAERHGTIGTQKDSATNLVQ 315  
QY 289 KIOTGRIDMOSVOTRLISDDPAKGLDRDRKRPPIAGRIODALKOPKESVYLVSPYFVPTK 348  
DB 316 LIPSG-----PGYTTEPNLR-----MENSIVHNAKERILICSPYFIPDE 354  
QY 349 SGTDLAKLYODGIDVYLTNSLOATDVAANVSGYKRPPLKAGITLYELOPHNAVPA 408  
DB 355 SLEAVTACYRGVTVLEFVS--EQADQFAIDHAOSSYQALLGAVKIYOF--PKPDV-- 409  
QY 409 TKDKGLTGSSVYSLAKTFIVD-----GKRIF--IGSFNLDPRSARLNTKMGVIESPKI 461  
DB 410 -----LHFKYMIADPPDPTGTGNEALGVLGSSNLDIRSGFLNTEISLMAKGNL 456



SQ Sequence 442 AA: 8.7%; Score 237; DB 22; Length 442;  
 Query Match 22.8%; Pred. No. 1.6e-13;  
 Best Local Similarity 58; Mismatches 156; Indels 88; Gaps 9;  
 Matches 89; Conservative 58; Mismatches 156; Indels 88; Gaps 9;  
 QY 76 IYLLNDPEAFARALLIESASHSLDLOYIWRNDISGRLLFNLYLAERGVRRLLID 135  
 Db 129 IDLFEDGHLXKVEVEDIYNADYIHLEXYFELDLGRIIDALETKIEGLLEVYLLD 188  
 QY 136 DNNTRGLDLDLALDHPNIEVRLNPFYLRKMRALGYLTD--FP-----RLNRRMH 185  
 Db 169 DVGSK-----KVRLSK--FKHFRALGCEVEFFPSKYPFLNFRMNNRH 230  
 QY 186 NKSFTADNRATILGGRNIGDEYFKVEDTVFADLDILATGSVYGEVSHDFDRYMASSAH 245  
 Db 231 KRIITIDGGIGYVGFNVGDDYLGLGKIGYWRDTHRYVGECIDALQLFLIDMNSQS-- 288  
 QY 246 NATRIIRSNICKGLQALGYNDETSHALLRREYVESPLYOKIOTGRIDMOSVOTRLI 305  
 Db 289 -----HRPQFKFDQKFFPKNG--DKGNALIQTA 315  
 QY 306 SDPAKGLDRDRKRPPIAGRLQDALKQPEKSYLVSPYFPTKSGDALAKLVODGIDVT 365  
 Db 316 SSGPAPDLHQ-----IEGYTKMISAKSKSYLOSPTIFIDOSTYNALKMANSSVEYN 369  
 QY 366 VLTNSLQATDAVAHSGYVYKRPPLKAGIKLYELOPHNAVPAKDKGLTGSSVTSLHAK 425  
 Db 370 LMIPC--KPDHPFYVWATFSNADLIDSGVNITYQNGF-----IHSK 410  
 QY 426 TFIYDGKRFIGSFNIDPRASRLNTEMGYVI 456  
 Db 411 IIMIDEISSIGSANMDFRSFELNEFVNAFI 441  
 RESULT 20  
 AAU57963  
 ID AAU57963 standard; Protein; 356 AA.  
 AC AAU57963;  
 DT 27-FEB-2002 (first entry)  
 DE Propionibacterium acnes immunogenic protein #18859.  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KN inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KM dermatological; osteopathic; neuroprotectant.  
 OS Propionibacterium acnes.  
 WO200181581-AZ.  
 PD 01-NOV-2001.  
 PF 20-APR-2001; 2001WO-US12865.  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 PA (CORI-) CORIAX CORP.  
 PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 DR WPT: 2001-616774/71.  
 DR N-PSDB; AAS59588.  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

XX Example 1; SEQ ID NO 19158; 1069pp; English.  
 PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 SQ Sequence 356 AA: 6.6%; Score 180; DB 22; Length 356;  
 Query Match 24.0%; Pred. No. 3e-08;  
 Best Local Similarity 40; Mismatches 102; Indels 86; Gaps 11;  
 Matches 72; Conservative 40; Mismatches 102; Indels 86; Gaps 11;  
 QY 84 EAFARALIESASHSLDLOYIWRNDISGRLLFNLYLAERGVRRV-----RLDLD 135  
 Db 104 DLYAAMIDIESAHOHTVYLETIFMKNDGEHFRQALVEAARGVYVAMDTFANLVVD 163  
 QY 136 DNNTRGLDLDLALDHPNIEVRLNPFYLRKMRALGYLTFPPLNRRMHKSFPA 191  
 Db 164 PRFFRQLE-----GVQVRA-QPLVTPSMITLRLG-----RDHKKLLIV 202  
 QY 192 DNRATILGGRNIGDEYFKVEDTVFADLDILATGSVYGEVSHDFDRYMASSAHNATRII 251  
 Db 203 DSKVAVYIGGYNIGSLYADWRDT-----HARITGAVGELFESVFDWM----- 245  
 QY 252 RSGNIGKGLQALGYNDETSHALL-RYRETVESPLYOKIOTGRIDMOSVOTRLSDPA 310  
 Db 246 -----NQRKGLIPRRNOPVLPD-----GVRYW-----DTA 273  
 QY 311 KGLDRDRKRP--PIAGRLQDALKQPEKSYLVSPYFPTKSGDALAKLVODGIDVT 367  
 Db 274 FAVHNSPMAVYPIRMNLTLEIDRASERIKIMWTGILLIPDDVVALQOASRGVDVRI 333  
 RESULT 21  
 ABG01152  
 ID ABG01152 standard; Protein; 280 AA.  
 AC ABG01152;  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #1143.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 WO200175067-AZ.  
 PD 11-OCT-2001.  
 PR 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB: AAS65339.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20: SEQ ID No 31511; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC (polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 280 AA:

Query Match 6.5%; Score 176; DB 22; Length 280;

Best Local Similarity 20.4%; Pred. No. 4.9e-08;

Matches 83; Conservative 44; Mismatches 122; Indels 158; Gaps 12;

QY 76 IYLLNDPHEFAARALISAESHSIDLYIWNDSIGRLLENLYLAERGVRLLLD 135

DB 11 IOLLENGEOYYPVFAKGAIGAOERILLETFTWEDVGQOLHALLAAORGKAVLLD 70

QY 136 DNNTRGLDILLALDSHPNIEVRLNPFVLRKRRALGYLDPRLNRMRNKSFTADNRA 195

DB 71 GYGSPLSDSEFVNELTAAAGVFRYIDP---RRLFGMRTN---VRRMRKTIIVIDARI 123

QY 196 TIILGNGIDGVEYKVEDIYFADLDILANGSVYGEVSHDDP-----RYMASHSA 244

DB 124 AFIIGGLNYSAEHMSYGPPEAKQDYAVRLSEPIVEDTL-QELENLPQGSAAARMWRH-- 180

QY 245 HNATRIIRSGNIGKGLQALGYNDSTRHALLARYRETVEGSPLOYKIQOTGRIDMQSVOTRL 304

DB 181 HKAEENRQPE---AQANY-----AVRF----- 202

QY 305 ISDDPAKGLDRRRKPPPIAGRLDALLKQPKSVYLVSPYFPTKSGTDLAKLVODGIDV 364

DB 203 -----GRL----- 205

QY 365 TVLTNSLGATDAVAHVGGYKRRPKLKAGIKLYELOPNNAVPATKDKGLTGSSVSLHA 424

DB 206 -----SGIT-----LVKGGVQVFEYRRR-----PLHG 227

QY 425 KTEIVDGKRIFIGSFNLDPRSARLNTMGVVISPKIAEQMERTLAD 471

DB 228 KVALMDHMTAVGSSNLDPLSLNLEANYIHD---RHFNQTLRD 270

RESULT 22

ABG01151

ID ABG01151 standard; Protein; 1097 AA.

XX AC ABG01151;

XX DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #1142.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB: AAS65338.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 20: SEQ ID No 31510; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC (polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1097 AA:

Query Match 6.5%; Score 176; DB 22; Length 1097;

Best Local Similarity 20.4%; Pred. No. 4.4e-07;

Matches 83; Conservative 44; Mismatches 122; Indels 158; Gaps 12;

QY 76 IYLLNDPHEFAARALISAESHSIDLYIWNDSIGRLLENLYLAERGVRLLLD 135

DB 241 IOLLENGEOYYPVFAKGAIGAOERILLETFTWEDVGQOLHALLAAORGKAVLLD 300

QY 136 DNNTRGLDILLALDSHPNIEVRLNPFVLRKRRALGYLDPRLNRMRNKSFTADNRA 195

DB 301 GYGSPLSDSEFVNELTAAAGVFRYIDP---RRLFGMRTN---VRRMRKTIIVIDARI 353

QY 196 TLIGRNIGDEYFKGEDTVFADLDILATGVSVDHPD-----RYWASHSA 244  
 DB 354 AFIGGNTVAEHHSSYGPAPKADYAVRLEGPIVEDIL-OFEELENLPGQSAARRMRRH-- 410  
 QY 245 HNAATRIISGNIGKGLQALGYNDETSHALLRYREVEESPLYQKIQOTRIMQSVQTL 304  
 DB 411 HNAENKRGGE---AQANGY-----AVVR----- 432  
 QY 305 ISDDPAKGLDRDRKRPPIAGRLQDALKQPEKSVYLVSPYFVPTKSGTDALAKLVQGDIV 364  
 DB 433 -----GRL----- 435  
 QY 365 TVLTNSLATDYAAVHSGVYKRPPLKAGIKLYELOPNHNAVPAKDKGLTGSSTSLHA 424  
 DB 436 -----SGYY-----LVKGVGVFEYRRR-----PLHG 457  
 QY 425 KFEIVDGKRIFIGSFNLDPRSARLNTENGCVIESPKIAEMERTLAD 471  
 DB 458 KVALMDHMAIVGSSNLDPLSLSLNEANVTIHD---RHFOQLRD 500

## RESULT 23

ABP28514  
 ABP28514 standard; Protein; 525 AA.

ABP28514;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 6204.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus pyogenes.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

Tetelin H;

WPI; 2002-352536/38.

N-PSDB; ABN69145.

New Streptococcus protein, for the treatment or prevention of infection  
 or disease caused by Streptococcus bacteria, such as meningitis, and  
 for detecting a compound that binds to the protein -

Claim 1; Page 3783; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B  
 streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 the specification. The proteins have antibacterial and antiinflammatory  
 activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
 antibodies that bind (I) are used in the manufacture of medicaments for  
 the treatment or prevention of infection or disease caused by  
 Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 Nucleic acids encoding (I) are used to detect Streptococcus in a  
 biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.

Sequence 525 AA;

Query Match 6.1%; Score 165; DB 23; Length 525;

Best Local Similarity 20.3%; Pred. No. 1.5e-06;

Matches 85; Conservative 58; Mismatches 123; Indels 152; Gaps 19;

QY 96 AEHSIDLQYI-----WFNDISGRLEFNLVLAERGVAVRLDDNNTRGLDPL-LIA 148  
 DB 185 AKKIYFLEFFLIAEQMW-----GELI-SLEKKVSEGEVAVLFD-----GMNLSLTS 233  
 QY 149 LD-----SHPNIEVRLF--NPFVLKRWALGYLTFPPLNRHMNKSFTADNRATILGG 200  
 DB 234 SDYARRLEQIGIKAKSFLPISPISTY-----NYRDRKIVLIDGEVSEFTGG 281  
 QY 201 RNIGGEYF-KVGEDTVFADLDILATGVS-----VGEVSHDFRYWASHSAHN 246  
 DB 282 INLADEYINVEYRFGHMDAGIMLEGEATDSFLILFLOMSTITEKELIIDPLSDHSLK- 340  
 QY 247 ATRIIRSGNIGKGLQALGYNDETSHALLRYREVEESPLYQKIQOTRIMQSVQTRLIS 306  
 DB 341 -----LPSDG-----VLPY-----G 351  
 QY 307 DDPAGGLDRDRKRPPIAGR--LDALKQPEKSVYLVSPYFVPTKSGTDALAKLVQGDIV 364  
 DB 332 DSP---LPTDK-----IGKNVYIDILNNAKEVYVMTFTYLLIDSEMERALRPAERGVDI 403  
 QY 365 TVLTNSLATDYAAVHSGVYKRPPLKAGIKLYELOPNHNAVPAKDKGLTGSSTSLHA 424  
 DB 404 RIMPV--PDKGVYALAKTYKALMSSGVKIVYEQGF-----VHS 444  
 QY 425 KFEIVDGKRIFIGSFNLDPRSARLNTENGCVIESPKIAEMERTLADTSPEYAVRYTL 482  
 DB 445 KVFISDNTRKAVVGTINLDYRSLYHFECAAT-----YLRVSV 481

## RESULT 24

ABP28513  
 ABP28513 standard; Protein; 531 AA.

ABP28513;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 6202.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 group A streptococcus; Streptococcus pyogenes; antibacterial;  
 antiinflammatory; infection; vaccine; meningitis; gene therapy.

Streptococcus agalactiae.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

PI Tettelin H;  
 XX  
 DR MPI: 2002-352536/38.  
 DR N-PSDB; ABN69144.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 PS Claim 1; Page 3783; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 SO Sequence 531 AA;  
 Query Match 5.9%; Score 160.5; DB 23; Length 531;  
 Best Local Similarity 19.8%; Pred. No. 4.1e-06;  
 Matches 82; Conservative 67; Mismatches 165; Indels 101; Gaps 14;  
 QY 74 SDIYLLNDPHEAFARAALIESAHSIDLOYIWRNDISRLFNLYLAEGRYRL 133  
 DB 169 TDVTFLEPTGEAFPEKMEELLKAKKYIFLEFFIIDEIGIMGELSTLEQVEEGVEYRIL 228  
 QY 134 LDD-----NNTRGDLDLLALDHPNIEVRLFPVLRKRWALGYLTDFPRLRRM 184  
 DB 229 YDGMIEITKLSFDYTKRLKLEI-----GIKAKAFSP-----SPFISTY--YVYRD 271  
 QY 185 HNKSFADNRATILGGRNIGDEYFKVGEDVFA---DLDLATGVSVEGVSHDPRYMAS 241  
 DB 272 HRKIVIVDVGVMGTGVNLADVEY--INHIEIFGMMKDSGIMLKRAVDSEFLFLQWMSI 329  
 QY 242 HSAHNATRIIRSGNIGLOA--LGYNDETSRHA--LLRREYEQSPLYOKIOTGRIDW 297  
 DB 330 TEE-----KMLVAPYLGVDLLENMGYIYP----- 356  
 QY 298 QSVOTRLISDDPAKGLDRDRRRPPIAGRLQDALKQPEKSVYLVSPVPTKSGTDALAKL 357  
 DB 357 -----GDSF---LDTDKVGENV---YIDILNHAKEYIYIMPPYIILDSLEHALQFA 402  
 QY 358 VODGIDVTVLNLSQATVAHVHSGYVYKRPPLKAGIKYLELOPNHAFATYDKGLTGS 417  
 DB 403 AERGVDVRIIMPGL--PDKPIPYALAKTYOALTKSGVKIYEX----- 443  
 QY 418 SVTSLAKTIVDGKRLIFGSFNLDPRSARLNTMGVYIESPKIAEOMERTLADT 472  
 DB 444 TLGFVHSKIFLSDNTKAVVGTINDYRSLYHNFECAYLYKVAIDIODIYRDYMDT 498  
 RESULT 25  
 ABP29889  
 ID ABP29889 standard; Protein; 531 AA.  
 XX  
 AC ABP29889;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 8954.

XX  
 KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB04789.  
 XX  
 PR 27-OCT-2000; 2000GB-0026333.  
 PR 24-NOV-2000; 2000GB-0028727.  
 PR 07-MAR-2001; 2001GB-0005640.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 PA (GENO-) INSY GENOMIC RES.  
 XX  
 PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR MPI: 2002-352536/38.  
 DR N-PSDB; ABN70520.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 PS Claim 1; Page 4016; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 SO Sequence 531 AA;  
 Query Match 5.9%; Score 160.5; DB 23; Length 531;  
 Best Local Similarity 19.8%; Pred. No. 4.1e-06;  
 Matches 82; Conservative 67; Mismatches 165; Indels 101; Gaps 14;  
 QY 74 SDIYLLNDPHEAFARAALIESAHSIDLOYIWRNDISRLFNLYLAEGRYRL 133  
 DB 169 TDVTFLEPTGEAFPEKMEELLKAKKYIFLEFFIIDEIGIMGELSTLEQVEEGVEYRIL 228  
 QY 134 LDD-----NNTRGDLDLLALDHPNIEVRLFPVLRKRWALGYLTDFPRLRRM 184  
 DB 229 YDGMIEITKLSFDYTKRLKLEI-----GIKAKAFSP-----SPFISTY--YVYRD 271  
 QY 185 HNKSFADNRATILGGRNIGDEYFKVGEDVFA---DLDLATGVSVEGVSHDPRYMAS 241  
 DB 272 HRKIVIVDVGVMGTGVNLADVEY--INHIEIFGMMKDSGIMLKRAVDSEFLFLQWMSI 329  
 QY 242 HSAHNATRIIRSGNIGLOA--LGYNDETSRHA--LLRREYEQSPLYOKIOTGRIDW 297  
 DB 330 TEE-----KMLVAPYLGVDLLENMGYIYP----- 356  
 QY 298 QSVOTRLISDDPAKGLDRDRRRPPIAGRLQDALKQPEKSVYLVSPVPTKSGTDALAKL 357

```

DB 357 -----GDS- -LDTRKVENV- -YIDILNHAREVYIMTPYLLDSELEHAIOFA 402
OY 358 VODGIDVLTNSLQATDVAAVHSGYVKYRKPLKAGIKLEYELOPHNAVPAATKDKLTGS 417
DB 403 ARGGVAVRIMPGI- -PDRIPYALAKTYQALTKSGVATIEY- - - - - 443
OY 418 SVTSLHAKTFIYDGRKIFIGSFNLDPRSARLNTMGVATIESPKIAQOMERTLADT 472
DB 444 TLGFVHSKIPLSDNTRKAVGTINLDYRSLYHFECAVAYLYKDAIODIYRDYMDT 498

RESULT 26
AAB70251
ID AAB70251 standard; Protein; 933 AA.
XX
AC AAB70251;
XX
DT 08-MAY-2001 (first entry)
XX
DE Human phospholipase D2 protein.
XX
KW Phospholipase D2; PLD2; autoimmune; rheumatoid arthritis; diabetes;
XX tumour; asthma; multiple sclerosis; stroke; HIV; T-cell activation.
XX
OS Homo sapiens.
XX
PN US6187559-B1.
XX
PD 13-FEB-2001.
XX
PF 28-AUG-1998; 98US-0141206.
XX
PR 28-AUG-1997; 97US-0057802.
XX
PA (NOVS ) NOVARTIS AG.
XX
PI Steed PM, Lasala DJ;
XX
DR WPI: 2001-190980/19.
XX
PT Isolated polynucleotide encoding human phospholipase D2 protein, useful
XX for treating autoimmune diseases, cancers, inflammatory diseases and
XX other diseases caused by altered phospholipase D activity -
XX
PS Claim 1; Column 13-14; 28pp; English.
XX
CC The present invention relates to phospholipase D2 (PLD2).
XX CC The invention is useful for producing phospholipase D polypeptide.
XX CC PLD2 polypeptide is useful for treating autoimmune diseases such as
XX rheumatoid arthritis and insulin dependent diabetes mellitus, tumors,
XX acute allergic reactions including asthma and chronic inflammatory
XX diseases, multiple sclerosis, atherosclerosis, restenosis, stroke and
XX coronary artery disease, activation of HIV (human immunodeficiency
XX virus) virus infection and associated HIV-dementia and T-cell
XX activation, and other PLD2 polypeptide associated diseases.
XX
SQ Sequence 933 AA:
Query Match 5.4%; Score 145.5; DB 22; Length 933;
Best Local Similarity 18.9%; Pred. No. 0.00028;
Matches 125; Conservative 85; Mismatches 175; Indels 275; Gaps 33;
OY 30 LICS- - - - -WLPPLERFESRHFNTSKPYRLDNIQT-RHYPH- - - - -T 70
DB 289 LILKSSYRQARMWAOETIELAOG- - - - -PGR- - -NFIQLHRHDSYAPPRGNLARWV 339
OY 71 NGLSDIYLLNDPHEAFARALIESAHS- - - - -LDIOYI- - - - -WRNDISGL 115
DB 340 NCAS- - - - -YFAVAADAILRAQIEFIITDMWLSPEVYLKRPASDDRLDI- - - - - 385
OY 116 LFNLYLAERGVAVRLLDNNTRGLDLDLLALDS- - - - -HPNIEVRLFNPEVL 165

```

```

DB 386 ---MFRKAAEEGVRSILLEKE- - - - -VELALGINSYKRALMLLHPNIKY-MRRPDV 436
OY 166 RKRVALGYLTDPRLRRMHNKSEFTADNRATILG- - - - -RINGD- - - - - 205
DB 437 TLM- - - - -AHHEKILVVDVAVAFGLGDLAYGRWDLHYRLTLGDSSEA 482
OY 206 - - - - -EYFKVED- - - - -TFEAD 218
DB 483 ASOLPTPRDSPAIPDFHNOFFMLGKDYSNLTKDWOLDRPFEDFIDRETTPRMPWD 542
OY 219 LDILATGSVVGESHDFDYMASHAHNATRIIRSGNICKGIALGYNDETSRHALLR 278
DB 543 VGVVHGLPARDLAREFIQW- - - - -NFKYTRA- - - - -KYKIPTYPLPKFT 586
OY 279 ETVEOSPLYOKIOTGRIDMOSVOTRLISDDPAKGLDRDRKPPAGRLDA- - - - -L 330
DB 587 STANOPF- - - - -TLPGGC- - - - -TVQV- - - - -LRSYDRWS- - - - -AGTLENSTLNNVLIHTI 630
OY 331 KQPKSVYLVSPFVFPFKSGTALAKLVODGID- - - - -VTVLT- - - - - 368
DB 631 RESQHFLYENOFIFISCSGDRVLNKGDEIVDRILKAHKQGCYRVYVLLPLPGFEGD 690
OY 369 - - - - -NSLQATDVAAVHSGY- - - - - - - - - - -VYRRLKAGIKIYELOPN 403
DB 691 ISTGGNSIQ- - - - -ALHFTYRTLCRGESILRLKAMGTWRDYISICGLRTGELG 746
OY 404 HAVPATRKDKGLTSSVTSLSHAKTFIYDGRKIFIGSFNLDPRS- - - - -ARLNTMGVATIE- - - - - 457
DB 747 HPV- - - - -SELTIHYSKVLADDRTYIIGSANINDRSLGKRSEILAVLIEDFT 796
OY 458 SPKI- - - - -AQOMERTLADTSPEYAYRYTLDNRNR- - - - -LQWHDPATRKTTPREPAAKMKRA 512
DB 797 EPSLMNGAEYQAGRPAISLRKHCFVILGANTRPDLDRPICDFF- - - - -QLMODMA 850

RESULT 27
AAB90563
ID AAB90563 standard; Protein; 476 AA.
XX
AC AAB90563;
XX
DT 29-JUL-2002 (first entry)
XX
DE Chlamydia pneumoniae cp6717 protein, SEQ ID NO:75.
XX
KW Chlamydial infection; antigen; immunogen; vaccine; diagnosis;
XX human respiratory disease; cardiovascular disease; atherosclerosis;
XX coronary artery disease; carotid artery stenosis; myocardial infarction;
XX cerebrovascular disease; aortic aneurysm; claudication; stroke;
XX strain CWL029.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FH Peptide 1..26
FH FT /label= signal_peptide
FT Protein 27..476
FT /note= "Mature protein"

MO200202606-A2.
PN 10-JAN-2002.
XX
XX 03-JUL-2001; 2001WO-IB01445.
XX
XX 03-JUL-2000; 2000GB-0016363.
XX
XX 11-JUL-2000; 2000GB-0017047.
XX
XX 21-JUL-2000; 2000GB-0017983.
XX
XX 07-AUG-2000; 2000GB-0019368.
XX
XX 18-AUG-2000; 2000GB-0020440.
XX
XX 14-SEP-2000; 2000GB-0022583.
XX
XX 10-NOV-2000; 2000GB-0027549.
XX
XX 22-DEC-2000; 2000GB-0031706.

```



XX (CHIR-) CHIRON SPA.  
 PA  
 XX  
 PI Rati G, Grandi G;  
 XX  
 DR WPI, 2002-154726/20.  
 XX N-PSDB; ABL91221.  
 PT Novel Chlamydia pneumoniae protein useful in the manufacture of a  
 PT medicament for treatment or prevention of infection due to Chlamydia,  
 PT preferably Chlamydia pneumoniae, and for diagnostic purposes  
 XX  
 PS Claim 1, Page 77, 364pp; English.  
 XX  
 CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia  
 CC pneumoniae (strain CMN029), and ABL91184-ABL91373 represent DNA encoding  
 CC them. The proteins are predicted to be immunogenic and may therefore be  
 CC useful in vaccine production and for diagnostic purposes. Chlamydia  
 CC pneumoniae is a common cause of respiratory disease in humans, and is  
 CC also involved in the development of cardiovascular diseases such as  
 CC atherosclerosis, coronary artery disease, carotid artery stenosis,  
 CC myocardial infarction, cerebrovascular disease, aortic aneurysm,  
 CC claudication and stroke. The proteins and nucleic acids of the invention  
 CC may be used in vaccines and pharmaceutical compositions for the  
 CC prevention or treatment of chlamydial infections, particularly Chlamydia  
 CC pneumoniae infections. The proteins may also be used in the detection of  
 CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched  
 CC DNA probe assay or blotting techniques for determining Chlamydia  
 CC pneumoniae gene expression. The present sequence represents a  
 CC specifically claimed Chlamydia pneumoniae protein of the invention.  
 XX  
 SQ Sequence 476 AA;  
 XX  
 Query Match 5.3%; Score 142.5; DB 23; Length 476;  
 Best Local Similarity 22.8%; Pred. No. 0.00018;  
 Matches 87; Conservative 50; Mismatches 132; Indels 113; Gaps 19;  
 XX  
 QY 145 LLLAL-DSHPNIEVRLNPFVLRKRWALGYLTDFPRLNR-----RMHNSKSTADNRAT 196  
 DB 110 LKALKERHNP---REFYVE-----TGCPSSTSLANVIEHMKLSIIDCKYC 155  
 QY 197 ILGGRNI-----GDEYFE-----VGEDTVFADLDILANGSVVG-EVSHDEDR 237  
 DB 156 ILGGRNEEFMCYRDEVEPKVNDPRLFVSGVRRLPFRDODIMLRSTAFGLDREYHK 215  
 QY 238 YMA--SHSANNATRIIRSGNIGKGLAGYNDERSHALLRYRETVSOS-----PLYOKI 290  
 DB 216 QFMMDIYAHN-----WFFIDNEPQFAGACPLUTLEQAEETVFPGFDKH 259  
 QY 291 QTRIDMQSVOTRLISDDPAK-----GLDRRRKRPPIAGRLADALKEPSVYLVSPFV 345  
 DB 260 ED-----LVLDSSKIRIYLGGRPHDKQPNPYTGKYLIGCARSSVLAHYFI 308  
 QY 346 PTSGCDALAKLYOD-GIDVTVLNLSIQATVAVHSGY-----VKY-----RKPLK 392  
 DB 309 PKBELNALVDVSHNMGVHSLITNGCHELS-PAITGPYAMGNRINFPALLYKRYPLMK 367  
 QY 393 AGIKIVELDPNHAVPATKDKGLTGSSV-----TSLHAFTYVDGRIFIGSFNDPRS 445  
 DB 368 KWF-CEUKRYERY-----STYEPAINETOLHKCKMIIDEIFVIGSYNEGKS 415  
 QY 446 ARLNTENGVIESPRIAEOMER 467  
 DB 416 DAFDEYISIVIESPEVAKANK 437  
 XX  
 RESULT 28  
 ID AAO16977 standard; Protein; 540 AA.  
 XX AAO16977;  
 XX  
 DT 27-MAY-2002 (first entry)

XX S cinnamomeus protein.  
 DE  
 XX  
 KW Promoter; phospholipase D; protein production; cardiovascular disease.  
 XX  
 OS Streptomyces cinnamomeus.  
 XX  
 PN WO200212485-A1.  
 XX  
 PD 14-FEB-2002.  
 XX  
 XX 06-AUG-2001; 2001WO-JP06757.  
 PE  
 XX 07-AUG-2000; 2000JP-0239132.  
 PR  
 XX (NEWI-) NEW IND RES ORG.  
 PA  
 PI Fukuda H, Kondo A, Kuroda S, Tanizawa K, Tokuyama S;  
 XX  
 DR WPI: 2002-227153/28.  
 XX N-PSDB; AAL45173.  
 XX  
 PT Streptococcus cinnamomeus-originated promoter, for constructing a  
 PT vector and transformant to produce phospholipase D -  
 XX  
 PS Disclosure; Page 45-48; 52pp; Japanese.  
 CC The present invention relates to a promoter originated from  
 CC Streptomyces cinnamomeus, which is capable of increased the production  
 CC of a protein when ligated upstream of its coding sequence. The promoter  
 CC is useful for constructing a vector and a transformant for the efficient  
 CC production of e.g. highly valuable phospholipase D by secretion, which is  
 CC used for treating cardiovascular diseases. The present sequence is a  
 CC protein shown in the exemplification of the invention.  
 XX  
 SQ Sequence 540 AA;  
 XX  
 Query Match 5.0%; Score 137; DB 23; Length 540;  
 Best Local Similarity 23.2%; Pred. No. 0.00074;  
 Matches 86; Conservative 41; Mismatches 121; Indels 122; Gaps 18;  
 XX  
 QY 185 HNSKSTADNRATILGGRNI-GDEYFVGEDTVFADLDILATGSGVGEVSHDEFRV--- 239  
 DB 201 HSKLLVDCESAVTGGINSKKDYY--VDQHPVTDVLDLALTGPAASAGRYLDLTWTWC 258  
 QY 240 -----ASHSANNATRIIRSGNIGKGLAGYNDERSHALLRYRETVSOS----- 261  
 DB 259 QNKSNIASVWFPAASGGDCMATEKEDANPRPAGPTGNVPLVAVGLGIGKIDSPASTFRP 318  
 QY 262 -----ALGYNDERSHALLRYRETV--EOSPPLYOKIQTRIDMQSVOTRLISDDP 309  
 DB 319 QLSAPDTRKVVLLPKTND---RDYDYNPESALRALYASA--DRQIV-----ISQO- 368  
 QY 310 AKGLDRRRKRPPIAGRLADALKEPSVYLVSPFVPTKSGTALAKLYODGIDVTVLN 369  
 DB 369 ----DINATCPRIAR-----YDVRLY-----DILAKMAAGVAVRIVAS 403  
 QY 370 SLQATVAVAHSGY-----VKYRKPRLKAGIKIVELDPNHAVPATKDKGLTGS 417  
 DB 404 D-PANRGAVSGSGYSQIKSLAETSDTLNRLLALTKG--DOOKAKAMKSTYOLGTFRS 459  
 QY 418 SVT-----SLHAFTYVDGRIFIGSFNDPRSARLNTENGVIESPRIAEOMERT 468  
 DB 460 SASATWADGHPYLDLHNRLLAVDVSSAFYIGSKNLYPSMLQ---DFGYIVESPEAKOLEAK 516  
 QY 469 LADTSPEYAY 478  
 DB 517 LLD--PEWKF 524  
 XX  
 RESULT 29  
 ID AAB70252 standard; Protein; 922 AA.  
 XX AAB70252  
 XX

```

XX AAB70252;
AC 08-MAY-2001 (first entry)
XX Human phospholipase D2b protein.
DE Phospholipase D2; PLD2; autoimmune; rheumatoid arthritis; diabetes;
XX tumour; asthma; multiple sclerosis; stroke; HIV; T-cell activation.
XX Homo sapiens.
XX US6187559-B1.
XX 13-FEB-2001.
XX 28-AUG-1998; 98US-0141206.
XX 28-AUG-1997; 97US-0057802.
XX (NOVS) NOVARTIS AG.
XX Steed PM, Lasala DJ;
XX WPI; 2001-190980/19.
XX Isolated polynucleotide encoding human phospholipase D2 protein, useful
XX for treating autoimmune diseases, cancers, inflammatory diseases and
XX other diseases caused by altered phospholipase D activity.
XX Example 3: Column 31-34; 28pp; English.
XX The present invention relates to phospholipase D2 (PLD2).
XX The invention is useful for producing phospholipase D polypeptide.
XX PLD2 polypeptide is useful for treating autoimmune diseases such as
XX rheumatoid arthritis and insulin dependent diabetes mellitus, tumors,
XX acute allergic reactions including asthma and chronic inflammatory
XX diseases, multiple sclerosis, atherosclerosis, stenosis, stroke and
XX coronary artery disease, activation of HIV (human immunodeficiency
XX virus) virus infection and associated HIV-dementia and T-cell
XX activation, and other PLD2 polypeptide associated diseases.
XX Sequence 922 AA:
SQ
Query Match 5.0%; Score 136; DB 22; Length 922;
Best Local Similarity 18.3%; Pred. No. 0.0022;
Matches 118; Conservative 87; Mismatches 184; Indels 256; Gaps 30;
OY 30 LICSSS-----WLPPEERTSRHFTSKVRDLNLIQI-RH--TPFTNGLSIDYLL 79
DB 289 LILKCSYRQARHMAOEITELAOG-----PGR--DFQLRHDSYAPPRGTLARMFV 339
80 NDPHEAFARALISAHSIDLQYI-----WRNDISGHLNLYILAEE 125
340 NCAGYFAAADAAILRAOEIIFITDMLSPDYVLRPAHSDMDRLDI-----MFKRAE 392
OY 126 RGVRYRLLDNDNTGLDLDLALD-----SHPNTEVRLNFVLRKKRALGY 173
DB 393 EGVRSYILL-----FKLELALIGNSYKRALMLHPNIKV-WRHPDQVTLW----- 439
OY 174 LTFDRLNRHMKSFADNRATILGG-----RNTGD----- 205
DB 440 -----AHHEKLLVYDQVAFLGGDLAYGRMDLHRYLTLDGDSSESAASQPPTR 490
OY 206 -----EYFKVED-----TVFADLLIATGCS 226
DB 491 PDSPTPLSHNQFWMLGKDYSNLITKDWVOLDREFEFDIRETTPRMFWRDVGVAHGL 550
OY 227 VVGEVSHDFDRYAWASHAHNAFRIRSGNIGKGLAALGYNDSHALLRYEYEQSL 286
DB 551 PARDLARHFIQW-----NFKTKTKA-----RYKPTPYLPLPKSTSTANQJPF 594
OY 287 YOKIOTGRIDWOSVOTRLISDDPAKGLDRRRKPIAGRLADA-----LKQPEKSVY 338

```

```

DB 595 --TLPGGQC--TTVOY-----LRSVDKWS-----AGTLENSILNAYLTIRESQFLY 638
OY 339 LVSPEYFVPTKSGFDALAKLEVODCID-----TVVLT-----NS 370
DB 639 IENQFFISCSGDKRTVLNKODEIYDRIKANKOGKCYRYVLLPLPGREGDISTGGNS 698
OY 371 LQATDYAAVHSGY-----VKYRPLKAGIKLYELQPNHVPATKD 411
DB 699 IQ-----AIIHFTYRTLRCGEYSILHRLKAAAGTAMRDYISICGLFRHGELGHPV----- 749
OY 412 KGLTSSVTSIHAKTFTIVGKRFISFNLDPRS--ARLNTENGVIYESPKIAEQMERTL 469
DB 750 -----SELYIHSKVLADDRVTYIGSANINDKSLGKRDSLEVALIED-----TEPESL 800
OY 470 ADTSPEYAVRYVTLDRHNR--LQWHDPATRKTYNPEPEAKLMKRIA 512
DB 801 MNGAEYQAGSVILGANTRPDLRLDIPICDDF-----QLMDMA 839

```

```

RESULT 30
AA44716
ID AA44716 standard; Protein; 933 AA.
XX
XX AA44716;
AC 25-APR-2000 (first entry)
XX
XX Human Phosphatidylcholine Phospholipase D isoform, hPLD 2.2.
DE
XX
XX Phosphatidylcholine Phospholipase D; PCPLD; hPLD 2.2; human; isoform;
XX phospholipid; PL; phosphatidic acid; PA; screen; drug; signal cascade;
XX inflammatory response.
XX Homo sapiens.
XX WO200000635-A1.
XX 06-JAN-2000.
XX
XX 30-JUN-1998; 98WO-US13515.
XX
XX 30-JUN-1998; 98WO-US13515.
XX
XX (CELL-) CELL THERAPEUTICS INC.
XX Leung DW, Tompkins CK;
XX
XX WPI; 2000-170925/15.
XX N-PSDB; AA249970.
XX
XX Isoforms of phosphatidylcholine phospholipase used to screen drug
XX candidates for identifying compounds to modulate phosphatidylcholine
XX phospholipase activity.
XX Claim 7; Page 46-49; 62pp; English.
XX
XX The present sequence is the human phosphatidylcholine phospholipase D
XX (PCPLD) isoform hPLD 2.2, isolated from a human liver cDNA library. This
XX sequence is identical to hPLD 2.1, with only eight amino acid changes.
XX It catalyses the hydrolysis of phospholipids (PL) to phosphatidic acid
XX (PA). The PCPLD isoforms are used to screen for drug compounds that can
XX modulate PCPLD activity. Compounds capable of inhibiting PCPLD activity
XX may be useful to inhibit a signal cascade in an inflammatory response.
XX The polynucleotides can be used for recombinant expression of the
XX polypeptide.
XX Sequence 933 AA:
SQ
Query Match 4.9%; Score 133.5; DB 21; Length 933;
Best Local Similarity 18.2%; Pred. No. 0.0038;
Matches 118; Conservative 90; Mismatches 187; Indels 255; Gaps 31;

```



```

XX AC AAY36906;
XX DT 07-OCT-1999 (first entry)
XX DE Chlamydia trachomatis lipoprotein sequence.
XX KM Vaccine: eye disease; conventional trachoma; nonendemic trachoma;
XX KM paratrachoma; inclusion conjunctivitis; genital disease; perithenitis;
XX KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
XX KM bartholinitis; pneumonia; venereal lymphogranulomatosis.
XX OS Chlamydia trachomatis.
XX PN W09928475-A2.
XX PD 10-JUN-1999.
XX PF 27-NOV-1998; 98WO-1B01939.
XX PR 04-NOV-1998; 98US-0107077.
XX PR 28-NOV-1997; 97ER-0015041.
XX PR 17-DEC-1997; 97FR-0016034.
XX (GIST ) GENSET.
XX PI Griffats R;
XX DR WPI: 1999-371125/31.
XX PT Genome sequence of Chlamydia trachomatis
XX PS Disclosure: Page 769; 1755pp; English.
XX CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
XX CC of Chlamydia trachomatis (see AAY36754). The polypeptide can be used as
XX CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX CC can also be used to control growth of the microorganism. Chlamydia
XX CC trachomatis is responsible for a large number of diseases, e.g. eye
XX CC diseases such as conventional trachoma, nonendemic trachoma,
XX CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX CC perithenitis, bartholinitis; pneumonia; venereal lymphogranulomatosis.
XX CC and venereal lymphogranulomatosis. The polypeptides of the invention
XX CC may be of use in treating these diseases.
XX SQ Sequence 489 AA;
XX
Query Match 4.8%; Score 130.5; DB 20; Length 489;
Best Local Similarity 19.8%; Pred. No. 0.0026;
Matches 97; Conservative 94; Mismatches 195; Indels 105; Gaps 23;
61 ILQIRHTPTNGLSD-----IYLLNDPHEAFARALIESAHSIDLYYWRNDISGR 114
DB 31 VSLMGFAHTIAPGDKKAKVLIHNGEYEHLLAASSAKYVELCPCLAGSEILST 90
OY 115 LLEFNLVYLAEGVRRVRL-----DDNTRGLDLDLIALDSHPNIEVLEFNPVLRKW 168
DB 91 VLQREORMEVPAVSYLVPTGICDDNDRKMLKTL--QENYD--RFF--YLFSDW 142
OY 169 RALGTLTPRRLNRRMHNKSTFADNRATILGSRNIGDEYFKGCEYVADLIL-ATGSV 227
DB 143 PRYCNV-FPNVT-BSHTKLSTVDKGYIFIGSNLEDOCKSGD---VDLEVSOSPRAV 196
OY 228 VGEVSHDEDRYASHANATRIIRSGNIGKGLQALGYNDEFSRHALLRYRETVESPLY 287
DB 197 IGVGLRP-----SAMRDQVTVYSEYCALIKRPFCAHAL--WRPFOKMLWN 243
OY 288 OKIQGRIDWQSVQTRLLISDDPAK-----DNRKPPYA 323
DB 244 KRLD-----DFRGIDPILNISTEKARSSFCAMIEISLCAVSPDLKMHFIRSGDESNTIA 299
OY 324 GRLQDALQPEKSVYLVSYFYPTKSGDALKLVOD-GIDVTYVLTNSL--QATDVAAVH 380

```

```

DB 300 EEVRLINOAHISIRIAQMFIFVAKIYDLSMAACWDRGVEIYLVNGBTRDSPEITRSY 359
OY 381 S--GVYK-----RKPL-----KAGIKLYELQNHVNPATKDKGLGSSVTSIH 423
DB 360 AMGNRINFPPLFFGSPPLMERFLYSPSRASKFVSEFYVA-----NTOLH 406
OY 424 AKTFIVDGKRIFFGSPNLDPSARLNTENGAVIESPKRIEOMERTLADTSPAYARVTL 483
DB 407 KCMVLVDHILVIGSYNPFCKSSMDCDECIYIDS-KEAVSKAQVVEKDLRLSKSVTH 465
OY 484 RHNRLQWH-DP 493
DB 466 --DIIMWYFDP 474

RESULT 33
ID AAM53281 standard; Protein; 932 AA.
XX AC AAM53281;
XX DT 16-JUL-1998 (first entry)
XX DE Murine phospholipase D 2 (mPLD2) protein.
XX KM Phospholipase D; PLD; murine; PLD mediator; autoimmune disease;
XX KM rheumatoid arthritis; psoriasis; ulcerative colitis; cancer; mPLD2;
XX KM inflammatory disease; wound healing.
XX OS Mus sp.
XX FH Key Location/Qualifiers
XX FT Misc-difference 923
XX FT /note= "encoded by NAA"
XX PN W09810076-A2.
XX PD 12-MAR-1998.
XX PF 06-AUG-1997; 97WO-US13919.
XX PR 05-SEP-1996; 96US-0025469.
XX PA (ONYX-) ONYX PHARM INC.
XX PI Frohman MA, Morris AJ;
XX DR WPI: 1998-193626/17.
XX DR N-PSDB: AAY20872.
XX PT DNA sequence encoding phospholipase D - useful to identify
XX PS modulators to treat auto-immune and inflammatory diseases
XX
Claim 16; Pages 59-61; 70pp; English.
XX
This represents a murine phospholipase D (PLD) 2 (mPLD2) protein. This
XX can be used in a method for identifying mediators of PLD, which comprises
XX transfecting a cell line with an expression vector comprising nucleic
XX acid sequences encoding a PLD protein and culturing the cell line in
XX culture medium, where the PLD protein is expressed stably. An effective
XX amount of a compound sufficient to cause a detectable loss in the
XX catalytic activity of PLD is added to the culture medium, and the loss in
XX phosphatidylcholine specific PLD activity. The PLD polypeptides which are
XX perinuclear membrane associated require PI(4,5)P2 for in vitro activity
XX and is activated by at least 1 g protein. The PLD polypeptides that are
XX plasma membrane associated, activates cytoskeletal reorganization
XX pathways, require PI(4,5)P2 for in vitro activity and does not require
XX Rac1m cdc42, RhoA, PKC or ARF1 for activation. A PLD mediator can be used
XX to treat autoimmune or inflammatory diseases, specifically rheumatoid
XX arthritis, psoriasis and ulcerative colitis. The mediator can also be
XX used in wound healing and for treating cancer and other diseases

```



QY 468 TLADTSPEYAVRT-----IDRHNRLQWHDPAATKRTY 499  
 Db 392 AFKEVSDVLAERATLGLRIDAOIRLO---DNYRQRY 424  
 RESULT 35  
 AAM89200  
 ID AAM89200 standard; Protein; 933 AA.  
 AAM89200;  
 18-MAR-1999 (first entry)  
 Human phosphatidylcholine phospholipase D (PCPLD) enzyme.  
 Phosphatidylcholine phospholipase D; PCPLD; enzyme; human; inhibitor;  
 phosphatidylcholine; phosphatidic acid; inflammatory response.  
 Homo sapiens.  
 US5859222-A.  
 12-JAN-1999.  
 17-DEC-1996; 96US-0768147.  
 15-DEC-1995; 95US-0008768.  
 17-DEC-1996; 96US-0768147.  
 (CELL-) CELL THERAPEUTICS INC.  
 Leung DW, Tompkins CK;  
 WPI; 1999-120006/10.  
 N-PSDB; AAV99778.  
 New nucleic acid encoding human phosphatidylcholine phospholipase D  
 and its complements - used to screen for specific inhibitors and  
 activators, potentially useful as anti-inflammatory agents  
 Claim 1; Columns 19-26; 27pp; English.  
 This represents a human phosphatidylcholine phospholipase D (PCPLD)  
 enzyme. A recombinant PCPLD enzyme can be used to screen for specific  
 inhibitors/activators. PCPLD inhibitors prevent conversion of  
 phosphatidylcholine to phosphatidic acid (potentially useful for  
 inhibiting an inflammatory response) and to raise antibodies (also  
 useful therapeutically). Fragments of the PCPLD nucleic acid are useful  
 as probes to identify, locate and measure corresponding nucleic acid in  
 biological samples and as amplification primers.  
 Sequence 933 AA;  
 Query Match 4.5%; Score 121.5; DB 20; Length 933;  
 Best Local Similarity 17.8%; Pred. No. 0.053;  
 Matches 115; Conservative 90; Mismatches 194; Indels 247; Gaps 30;  
 QY 30 LLCSSS-----WLPPEERTSRHNTSKPVRDLNLIQI-RH---TPHNGLSDIYLL 79  
 Db 289 LILKSSSYRQAWMAQOEITELAQG-----PGR--DFQLRHDSYAPRRGTILARMFV 339  
 QY 80 NDPHAFAPARALIAEHSIDLQYIT-----WRNDISGRLEFNLYLAAE 125  
 Db 340 NGAGYFAAVVAAILRAOEIIFITDMWLSPEVYLKRPANSDWRDLDI-----MKRAAE 392  
 QY 126 RGVRYRLLDONNTRGLDLDLALDS-----HPNIEERLNPFLYRKWRALGYIT 175  
 Db 333 EGVRSILLFNE-----VELAGINSYISKALMLLHPVITKV-MKHPOQVTLW----- 439  
 QY 176 DEPRLRMRHNSKFTADNRATILIG-----RNIGD----- 205  
 Db 440 -----AHHEKLLVVDGVVAFGLGDLAYGRWMDLHYRDLGDSSESASQPTTPRPD 492

QY 206 -----EYKVGED-----TVFADDLILATGSVY 228  
 Db 493 SPATPDLISHNOEFWLGKQVSNLITNDWQVQDRPFEDFIDRETTTPRPMRDVGVVHGLPA 552  
 QY 229 GEVSHDFEDRYAMSHSAHNATRIIRSGNICKGLQALCYNDETSRHALIRREVEQSPLYQ 288  
 Db 553 RDLARHFIQRW-----NFTKTTKA-----KYKTPYPLLPKSTANDLPF-- 594  
 QY 289 KIOTGRIDQSVOTRLLISDDPAKGLDRDRKPPVIGRLQDA-----LKQPEKSVYLV 340  
 Db 595 TLPGGOC--TIVQV-----LRSYDRWS-----AGTLENSILNAYLHTRISQHPHYIE 640  
 QY 341 SPYFVPTKSGTDALAKLVODGID-----VTVLTNSLOATD-----V 376  
 Db 641 NOFFISCSGDRIVLKNKVGDEIVDLIKAKQGCWYRVYVLLPLPGFEGDISTGGKSIQ 700  
 QY 377 AAVHSGY-----VKYRKPLLAGITLYELQPNHAYPAIKDKGLIGS 417  
 Db 701 AILHFTYRTLGRGEYSILRLKAMGTAWRDYISIGLTFHELGHHPV-----S 750  
 QY 418 SVTSLHAKTFIVDGKRIFIGSFNDPRS--ARLNTMGVIE-----SPKI--AEQMERT 468  
 Db 751 ELIYHSKVLADDRVITIDSANINDRSLGKRDESLAVLIDTETEPSLMNGATEYQAGR 810  
 QY 469 LADTSPEYAVRTLDNRNR--LQWHDPAATKRTYPNEPEAKLWKRIA 512  
 Db 811 FALSIRKHCFSVILGANTRPDLDRPICDDPF-----QLWQDMA 850  
 RESULT 36  
 AAY44715  
 ID AAY44715 standard; Protein; 933 AA.  
 AAY44715;  
 25-APR-2000 (first entry)  
 Human phosphatidylcholine Phospholipase D isoform, hPLD 2.1.  
 Phosphatidylcholine Phospholipase D; PCPLD; hPLD 2.1; human; isoform;  
 phospholipid; PL; phosphatidic acid; PA; screen; drug; signal cascade;  
 inflammatory response.  
 Homo sapiens.  
 WO200000635-A1.  
 06-JAN-2000.  
 30-JUN-1998; 98WO-US13515.  
 30-JUN-1998; 98WO-US13515.  
 30-JUN-1998; 98WO-US13515.  
 (CELL-) CELL THERAPEUTICS INC.  
 Leung DW, Tompkins CK;  
 WPI; 2000-170925/15.  
 N-PSDB; AAZ49969.  
 Isoforms of phosphatidylcholine phospholipase used to screen drug  
 candidates for identifying compounds to modulate phosphatidylcholine  
 phospholipase activity -  
 Claim 6; Page 30-33; 62pp; English.  
 The present sequence is the human phosphatidylcholine phospholipase D  
 (PCPLD) isoform hPLD 2.1, isolated from a human lung cDNA library. It  
 catalyses the hydrolysis of phospholipids (PL) to phosphatidic acid  
 (PA). The PCPLD isoforms are used to screen for drug compounds that can  
 modulate PCPLD activity. Compounds capable of inhibiting PCPLD activity  
 may be useful to inhibit a signal cascade in an inflammatory response.

D6 751 ELIYIHSKVLADDRTVIIDSANINDRSLGKRDSLAVLIEDTETEPSLMNGAEYQAGR 810

QY 469 LADTSPEYAVRYTLDRHNR--LQWHDPAIRKTYPNEPEAKLMKRIA 512  
 DB 811 FALSIRKHCFSVILGANTRPDLRDLRDCDDEF-----QLQMDMA 850

RESULT 38

AA017567  
 ID AA017567 standard; Protein; 507 AA.

AA017567;

19-JUL-2002 (first entry)

M catarrhalis MCA100564 protein SEQ ID NO: 14.

Moraxella; vaccine; respiratory tract infection; antiinflammatory;  
 audiotory; antibacterial; otitis media; sinusitis; pneumonia.

Moraxella catarrhalis.

WO200218595-A2.

07-MAR-2002.

28-AUG-2001; 2001WO-CA01221.

28-AUG-2000; 2000US-228294P.

28-AUG-2000; 2000US-228295P.

28-AUG-2000; 2000US-228296P.

29-AUG-2000; 2000US-228438P.

29-AUG-2000; 2000US-228439P.

29-AUG-2000; 2000US-228440P.

29-AUG-2000; 2000US-228441P.

29-AUG-2000; 2000US-228442P.

29-AUG-2000; 2000US-228443P.

29-AUG-2000; 2000US-228511P.

29-AUG-2000; 2000US-228512P.

29-AUG-2000; 2000US-228742P.

29-AUG-2000; 2000US-228773P.

01-SEP-2000; 2000US-229465P.

01-SEP-2000; 2000US-229474P.

01-SEP-2000; 2000US-229475P.

01-SEP-2000; 2000US-229478P.

05-SEP-2000; 2000US-229740P.

05-SEP-2000; 2000US-229803P.

05-SEP-2000; 2000US-229804P.

05-SEP-2000; 2000US-229805P.

05-SEP-2000; 2000US-229806P.

05-SEP-2000; 2000US-229809P.

06-SEP-2000; 2000US-229811P.

06-SEP-2000; 2000US-230214P.

06-SEP-2000; 2000US-230250P.

(AVET ) AVENTIS PASTEUR LTD.

Loosmore S, Wang J, Bradley B, Ochs M, Yang Y;

WPI; 2002-401721/43.

N-PSDB; AAL46499.

Moraxella polypeptide and polynucleotides useful as vaccine for  
 immunizing a host e.g. humans against disease e.g. otitis media,  
 pneumonia, caused by infection of the bacteria

Claim 28; Fig 13; 277pp; English.

The present invention provides the protein and coding sequences of  
 CC proteins from Moraxella catarrhalis. These can be used to produce  
 CC vaccines which protect against M. catarrhalis infection, which can  
 CC cause otitis media, respiratory infection, sinusitis, and pneumonia. The  
 CC present sequence is a protein of the invention.

XX SQ Sequence 507 AA;

Query Match 4.3%; Score 118; DB 23; Length 507;  
 Best Local Similarity 21.3%; Pred No. 0.043; Mismatches 193; Indels 134; Gaps 27;  
 Matches 109; Conservative 76;

QY 24 ISLLCLLCSCSSWLPPLERTESRHFNTSKPRVDNIQIHTPTNGLSDIYLLDPH 83

DB 12 LSLALIGMACSS-IP-----KKIDTSAD-----LLAVNVP---MKGGYGVYDD 53

QY 84 EAFARAALIESAEHSLDQIYIWRNDISGRLLFNLYIAARGVAVLLDDNNTRGID 143

DB 54 TTSVAQAPSVASLR-----WQEFYTPRLALIELALD-----\*-NNKDIQ 92

QY 144 DLLALDSHPNIEVRLNFPVLRKWRALGYLDPFRLNRMHNSFT--ADNRATILGCR 201

DB 93 SAVLAVQSA-----RAOQITAEAGSVPOVGSNTSVTQANNRIIDANST 136

QY 202 NIGDEYFKVGEPTVPAADDILATGSGVGEVSHDEPRYMAASHANATRIIRSGNIGGLQ 261

DB 137 N-----YHVGILAMSGYEIDL--WGRVASQKQALHQYIATNAKDAVOISITSSVAQGVY 189

QY 262 ALGYNDETSRHALLRYRETVESQPLYOKIQIGRIDMQSVQTRLISDDPAKGLDRRRKP 321

DB 190 NLA-----HRLAQ-RQLAEQT-----LKTREHAMMITQKRF---EAGT--DSKSPS 229

QY 322 IAGRLQDALKQPEKSVYLVSPYFVPTKSGTDALAKLVODGIDVYLVLTNSLQATDVAAVHS 381

DB 230 L--QAASSLESARLAVYAADTSILKAKNALQL-----IGRPVNNELPAID-ASMHM 279

QY 362 GYVKYRKPLLRAGIR---LYE---LQPNHAVPA-----TKDKGLTG 416

DB 280 GHIT-TQTLFSGALPSELLYYRPDIQAEHRLKAGANINVARAAYFPSIRLSSVNGFSS 338

QY 417 SSVTSLHAKTFI--VDGKRIFGSEFNLPDSARLNTGCVTIESPKIAE---OMERTLAD 471

DB 339 NSLNLFESSALGWSFGRAISLPF--DAGSRRAHBEAQAQSAALVDEYKAIONAKRE 396

QY 472 TSPEYAVRYT---LDRHNRLOQWHDPAIRKTY 499

DB 397 VSDVLAERATFLGLRDAQIRLQ---DNYRQTY 425

RESULT 39

ABB92080  
 ID ABB92080 standard; Protein; 832 AA.

ABB92080;

31-MAY-2002 (first entry)

Herbicideally active polypeptide SEQ ID NO 1291.

Herbicideal; plant; agriculture; herbicide.

Arabidopsis thaliana.

WO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB ) BAYER AG.

Tietjen K, Weidner M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicideally active compounds,





Db 1180 LOSQWYOHNLNLCGYRV-----NAIKDLYL:---EBSFTDKENITNLITKL 1223  
QY 93 ISAEHSLD-----LQYIWRNDISGRLEFNLYLAERGVR-----RLLDNN 138  
Db 1224 VONEEHQNDSPARQMDKVKXN--KAALLIGLYLNRYGVKFGDVNIKELMLFKPD 1281  
QY 139 TRG-----LDDLLALDSHPNIE-VLEFNE--VLKWRALGYLTDPPRLNRMHNSFT 190  
Db 1282 FYGEKYSVLDRLIEIGSKENNIKSGSRFTFAFGCVLAKYTKSGNLDAFLNYNROL---FT 1337  
QY 191 ADNRATILGSRNIGDEYFVGEEDTVFADLILATGVSVEGSHD---FDRYWASHAHN 246  
Db 1338 NID-----NNDWFIDATEDHYT---IARASEVEEIKNSKHAPDNLKRSHURNT 1385  
QY 247 ATRIIRSGNIGKGLAL--GYNDETSRHALLRYRETVEO-SPLYOKIQGRIDWOSVQTR 303  
Db 1386 ILPLT---NIDKAHLXLYLSNYNAIAFGSABRLGKKSLEDIKDIYNKADGYRNYDQWYR 1442  
QY 304 LISDPAKGLDRDRKP-----PIAGRLODALKQPEKSYLL-VSPYFVP-----TKSGT 351  
Db 1443 LASDNKQRLRDAVPIWEGYNAPGWEKYGRYNTDKVYTPLEFFGPMKYYNYNGT 1502  
QY 352 DALAKLVODGIDVTYVLTNSLQATDYAAVHSGYVKYRRPPLKAGIKLYELOPNHAYPATKD 411  
Db 1503 GAYAAIYPNSDOI-----RTDVKYVHLEWVG-----EYGISVYTHETTH----- 1541  
QY 412 KGLTGSSVTSIAKFIYVDGKRIFIGSENLDPBSARLNTBMGVYIESPKIAEOMERTLA- 470  
Db 1542 -----VNDRAIYLGF-----GHREGT-----DAEAYAOGMLOTPTVY 1573  
QY 471 ---DTSPEYAVRVTLDRHN-RLQWH--DPATPKT 498  
Db 1574 GSGFDEFGSLGINMVFRRKNDGNQWYITDPKITKT 1608

Search completed: May 12, 2003, 09:57:21  
Job time : 88 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:55:41 ; Search time 46 Seconds  
(without alignments)  
1097.186 Million cell updates/sec

Title: US-10-066-551-4

Perfect score: 2713

Sequence: 1 MRANPKTQAMPSETISLTKT.....KLWKRIAKILSLPIESLL 525

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2713	100.0	525	2	B81859
2	2599	95.8	508	2	B81083
3	1115.5	41.1	493	2	C64847
4	1113.5	41.0	493	2	H90806
5	1106.5	40.8	493	2	D85666
6	1100	40.5	495	2	AE0636
7	821	30.3	518	2	G97556
8	821	30.3	518	2	AB3777
9	808.5	29.8	529	2	F82983
10	632	23.3	502	2	E71963
11	628	23.1	502	2	F64543
12	427	15.7	652	2	B83724
13	325	12.0	482	2	G70063
14	318	11.7	504	2	A11433
15	311	11.5	504	2	A11075
16	308.5	11.4	494	2	F90001
17	293.5	10.8	482	2	H86744
18	293.5	10.8	482	2	A11762
19	288.5	10.6	482	2	AG3387
20	281	10.4	476	2	T43863
21	275	10.1	505	2	G89906
22	274.5	10.1	490	2	B82971
23	274	10.1	503	2	B84007
24	273.5	10.1	503	2	C84125
25	270.5	10.0	500	2	S60089
26	264.5	9.7	492	2	AE3539
27	262.5	9.7	467	2	H82711
28	261	9.6	359	2	D83103
29	258	9.5	510	2	E97307

30	253	9.3	480	2	AH2333	cardiolipin synthase
31	248.5	9.2	484	2	C82171	cardiolipin synthase
32	239	8.8	486	2	AG0266	cardiolipin synthase
33	238.5	8.8	487	2	F86771	cardiolipin synthase
34	231	8.5	398	2	S55419	cardiolipin synthase
35	231	8.5	401	2	B83377	cardiolipin synthase
36	229	8.4	413	2	A10598	probable phospholipase
37	227	8.4	492	2	B97658	cardiolipin synthase
38	227	8.4	492	2	AC2882	cardiolipin synthase
39	222.5	8.2	532	2	G82872	cardiolipin synthase
40	222	8.2	413	2	C90737	cardiolipin synthase
41	222	8.2	413	2	D85587	cardiolipin synthase
42	222	8.2	413	2	E64815	probable synthetase
43	219	8.1	474	2	B95023	cardiolipin synthase
44	219	8.1	510	2	D97894	cardiolipin synthase
45	216	8.0	486	2	AF0651	cardiolipin synthase

#### ALIGNMENTS

##### RESULT 1

B81859  
phospholipase D-family protein NMA1646 [imported] - Neisseria meningitidis (strain Z24)  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: B81859  
R:Parthill, J., Achtman, M., James, K.D., Bentley, S.D., Churcher, C., Klee, S.R., Moore, H., Holtz, S., Jørgensen, K., Leather, S., Mouton, S., Mungall, K., Quail, M.A., Rajanore, Nature 404, 502-506, 2000.  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: AB1775; MUID:20222556; PMID:10761919  
A:Accession: B81859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-525 <PAR>  
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84874.1; PID:g738  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1646

Query Match	100.0%	Score 2713	DB 2	Length 525
Best Local Similarity	100.0%	Pred. No. 1.3e-191		
Matches 525	Conservative	0	Mismatches 0	Indels 0
			Gaps 0	
QY	1	MRANPKTQAMPSETISLTKRSLISLCLLSCSSMLPPLERTESRHNTPKPYRLDN	60	
DB	1	MRANPKTQAMPSETISLTKRSLISLCLLSCSSMLPPLERTESRHNTPKPYRLDN	60	
QY	61	ILQIRTPHTNGSLDIYLLNDPHEAFARALIESAHSIDLYQYIWRNDISGRLEFLNY	120	
DB	61	ILQIRTPHTNGSLDIYLLNDPHEAFARALIESAHSIDLYQYIWRNDISGRLEFLNY	120	
QY	121	YLAERGVRRLLDNNTRGIDLLDLDLSDHPNIEVRLFPVPLKRWALGYLTFPRL	180	
DB	121	YLAERGVRRLLDNNTRGIDLLDLDLSDHPNIEVRLFPVPLKRWALGYLTFPRL	180	
QY	181	NRMHNKSFADNRATILGGRNIGDEYFKVGEPTVADIDILATGSVGVSHDFRYNA	240	
DB	181	NRMHNKSFADNRATILGGRNIGDEYFKVGEPTVADIDILATGSVGVSHDFRYNA	240	
QY	241	SHAHNATRIIRSGNIGKGLGYNDERSHALRYRETVESPLYOKIOTGRIMOSV	300	
DB	241	SHAHNATRIIRSGNIGKGLGYNDERSHALRYRETVESPLYOKIOTGRIMOSV	300	
QY	301	QTRLISDDPAKGLDRDRRKPPIAGRLDALKQPEKSVYLVSPVFTKSGTDLAKLYOD	360	
DB	301	QTRLISDDPAKGLDRDRRKPPIAGRLDALKQPEKSVYLVSPVFTKSGTDLAKLYOD	360	
QY	361	GIDVTYVLTNSLOATDVAAVHSGVYKRRPLKAGICILYELQPHNAPATKDKGLTSSVT	420	
DB	361	GIDVTYVLTNSLOATDVAAVHSGVYKRRPLKAGICILYELQPHNAPATKDKGLTSSVT	420	

QY 421 SLHAKTFVDCRKRIFGSPNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVR 480  
 |||||||  
 Db 421 SLHAKTFVDCRKRIFGSPNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVR 480  
 QY 481 TLDRNRLOWHDPATRKTYPNPEPAKLMKRIAAILSLPIESLL 525  
 |||||||  
 Db 481 TLDRNRLOWHDPATRKTYPNPEPAKLMKRIAAILSLPIESLL 525

# RESULT 2

B81083

cardiolipin synthetase family protein NMB1434 [imported] - Neisseria meningitidis (strain C:Species: Neisseria meningitidis

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001

C:Accession: B81083

R:Jettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.R.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

rt, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizze, M.

Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755; PMID:10710307

A:Accession: B81083

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-508 <TET>

A:Cross-references: GB:AE002494; GB:AE002098; NID:97226684; PIDN:AAF1795.1; PID:9722667

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB1434

Query Match 95.8%; Score 2599; DB 2; Length 508;  
 Best Local Similarity 96.8%; Pred. No. 3,1e-183;  
 Matches 502; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 18 MKTSLISLCLLSCSCSMPLPEERTSRHFNTPKRVRLNDIQLRHPHTNGLSDIY 77  
 |||||||  
 Db 1 MKTSLISLCLLSCSCSMPLPEERTSRHFNTPKRVRLNDIQLRHPHTNGLSDIY 77  
 QY 78 LLNDPHEAFARALIESAEHSLDQYIWRNDISGRLLFNLVYLAERGVRRLLDN 137  
 |||||||  
 Db 61 LLNDPHEAFARALIESAEHSLDQYIWRNDISGRLLFNLVYLAERGVRRLLDN 120  
 QY 138 NTRGLDILLALDSHPNIEVRLNPEVLRKRAIGYLTDFPLNRHMKSFADNRATI 197  
 |||||||  
 Db 121 NTRGLDILLALDSHPNIEVRLNPEVLRKRAIGYLTDFPLNRHMKSFADNRATI 180  
 QY 198 LGRNIGDEYKVGEDYFADLDILATGVSVEVSHDFRWASHANATRIIRSGNIG 257  
 |||||||  
 Db 181 LGRNIGDEYKVGEDYFADLDILATGVSVEVSHDFRWASHANATRIIRSGNIG 240  
 QY 258 KGLALGYNDETSRHALLRYRETVESQSPLYOKIOTGRIDMOSVOTRLISDDPAKGLDR 317  
 |||||||  
 Db 241 KGLALGYNDETSRHALLRYRETVESQSPLYOKIOTGRIDMOSVOTRLISDDPAKGLDR 300  
 QY 318 RKPPLAGLQALQOPEKSVYLVSPYFPTKSGDALAKLVQDGDVTVLNSLCATVA 377  
 |||||||  
 Db 301 RKPPLAGLQALQOPEKSVYLVSPYFPTKSGDALAKLVQDGDVTVLNSLCATVA 360  
 QY 378 AVHSGYVYRRPLKAGIKYELQPNHAPATKDKGLGSSVTSLHAKTFVDCRKRIFG 437  
 |||||||  
 Db 361 AVHSGYVYRRPLKAGIKYELQPNHAPATKDKGLGSSVTSLHAKTFVDCRKRIFG 420  
 QY 438 SFNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVRVLTLDRNRLOWHDPATRK 497  
 |||||||  
 Db 421 SFNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVRVLTLDRNRLOWHDPATRK 480  
 QY 498 TYPNEPEAKLMKRIAAILSLPIESLL 525  
 |||||||  
 Db 481 TYPNEPEAKLMKRIAAILSLPIESLL 508

# RESULT 3

C64847  
 ymdc protein - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C:Accession: C64847

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: C64847

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-493 <BIAT>

A:Cross-references: GB:AE000206; GB:U00096; NID:91787282; PIDN:AACT4130.1; PID:917872

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: ymdc

Query Match 41.1%; Score 1115.5; DB 2; Length 493;

Best Local Similarity 48.3%; Pred. No. 3.7e-74;

Matches 224; Conservative 81; Mismatches 142; Indels 17; Gaps 5;

QY 72 GLSDIYLNDPHEAFARALIESAEHSLDQYIWRNDISGRLLFNLVYLAERGVR 131  
 |||||||  
 Db 37 GOCGLFPLEKSLDAFAARYRLAEAEHSLDQYIWRNDISGRLLFNLVYLAERGVR 96  
 QY 132 LLDDNNTRGLDILLALDSHPNIEVRLNPEVLRKRAIGYLTDFPLNRHMKSFETA 191  
 |||||||  
 Db 97 LLDDNNTRGLDILLALDSHPNIEVRLNPEVLRKRAIGYLTDFPLNRHMKSFETA 156  
 QY 192 DNRAITLGRNIGDEYKVGEDYFADLDILATGVSVEVSHDFRWASHANATRII 251  
 |||||||  
 Db 157 DGVTLVGGRNIGDEYKVGEDYFADLDILATGVSVEVSHDFRWASHANATRII 216  
 QY 252 RSGNIGK-----LQALGYNDETSRHALLRYRETVESQSPLYOKIOTGRIDMOSVOTRL 304  
 |||||||  
 Db 217 ---DYPEGEAMDRLELPSMINDAMTH---RYLRKMSSPFIMHVLGDITPLIMAKTRL 269  
 QY 305 ISDDPAKGLDRRKPPIAGRLQALQOPEKSVYLVSPYFPTKSGDALAKLVQDGDV 364  
 |||||||  
 Db 270 LSDDPAKGEKGRKRRSLLPQRLFPIMGSPSERIDIISSYFPTKAGVQALIRMKRVKI 329  
 QY 365 TVLNSLCATVAAVHSGYVYRRPLKAGIKYELQPNHAPAT--KDKGLGSSVTS 423  
 |||||||  
 Db 330 AILNLSLAANAAVAVHAGYARWRKRLRYGVELKELPTEQSSLTDRGTTGSGASLH 389  
 QY 424 AKTFVDCRKRIFGSPNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVRVLTD 483  
 |||||||  
 Db 390 AKTFVDCRKRIFGSPNLDPSARLNTMGVIESPKIAEQMERLADTSPEYAVRVLTD 449  
 QY 484 RHNRLQWHDPAATR--TYPNEPEAKLMKRIAAILSLPIESLL 525  
 |||||||  
 Db 450 RMGRINWVDRHAKKEIILKKEPATSPMKRVAVRSLIPVWML 493

# RESULT 4

H90806

probable synthase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: H90806

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasawara, N.; Yasunaga, T.; Kuhnara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9

A:Reference number: A96629; MUID:21156231; PMID:11258796

A:Accession: H90806

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA034847.1; PID:913360888; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:  
A;Gene: ECS1424

Query Match	41.0%;	Score 1113.5;	DB 2;	Length 493;
Best Local Similarity	-48.1%;	Pred. No. 5.2e-74;		
Matches 223;	Conservative 82;	Mismatches 142;	Indels 17;	Gaps 5;

```

QY 72 GLSDIYLLNDHDEHFAAARALIESAHSILDOXYITWRNDISGRLLFNPLYLAERGVVR 131
Db 37 GOCGLFPLEKSLDFAARYRLAEMSERTLDQYITWODMSGRLLFSLLAARKGVVR 96
QY 132 LLLDDNNTRGLDLDLLADSHPNIEVRLNFPVLRKWRALGYLTDFPLNRHMKSEFA 191,
Db 97 LLLDDNNTPGLDILRLDLSHPRIEVRLENFPFSRLLRPLGITYDFESLRNHRMKSPFV 156
QY 192 DNKATILGGRIGDEYFVGEDYVFPADILLATGSSVYGEVSHDPRWYASAHNAATRII 251
Db 157 DGVYTLVWGRIIGAYFAGEEPLEFSDLVMAICPVEDVADDFARYWYCKSVSPLOOVL 216
QY 252 RSGNIGK-----LQALGYNDETSRHALLREYEVOSPLQYKICGTGRIDMGSVOGR 304
Db 217 ---DYPEEDMADRIELPASHMINDAMTH----RYLRKMSSPFILNHLVDGTLPIIMAKTRL 269
QY 305 ISDPPAKGLDHRDRKRPPIAGRLQDALQOPEKSVYLVSPFVPTKSGTDALAKLVQDGDIV 364
Db 270 LSDPPAKEGCKARKSHLLPQRLFDIMGSPSERIDIISSYFVPTRAGVQQLDRMRKGVKI 329
QY 365 TVLTPNSIQATPVAAVHSGVYVYKRRPLKAGIKLYELOQNAHVAPT-KDKGLTGSSVYSLH 423
Db 330 AILTPNSLANVAVYAHAYARWKRKLLRYGVELLEKATREQSSLTLDHRGITGNSGASLH 389
QY 424 AKTFIVDCKRIFIGSFNLDPRSARLNTEMGVIVISPKIAEOMERTLADTSPZYAVRYVL 483
Db 390 AKTFSIDCKYVFIGSFNDRPSTLNTTEMGVISESETLAQILDRFIQSOYDAAWQRLD 449
QY 484 RHNRLQWHDPRTRK--TYPNPEPAKLKRIAKILSLPISSL 525
Db 450 RMGRILNVYDRHAKREIYLKKEPATSPFMKRVYVRLASILPFWML 493

```

```

RESULT 5
D85666
Probable synthase ymdc [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: D85666
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.N.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206531
A:Accession: D85666
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <STO>
A:Cross-references: GB:AE005174; NID:g12514578; PIDN:AAG55792.1; GSPDB:GN00145; UMGP:Z16
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: ymdc

```

	Query Match	Similarity	40.8%;	Score	1106.5;	DB	2;	Length	493;	
	Best Local	Similarity	47.8%;	Pred.	No.	1.7e-73;				
	Matches	222;	Conservative	82;	Mismatches	143;	Indels	17;	Gaps	5
Oy	72	GLSDIYLLNDPDEEAAARALAESAEHSIDLYYYIIRNDISGRLLFNLYLAAERGVRR	131							
	:: :		:  :     :	:	:					
Db	37	GCGGFLPELEKSIDAFRAAKRLAEMSHITDVOYTIQDDMSGRLFSALLAAAKRGVRR	96							
Oy	132	LLLDNNFTGCLDLLDLALDSHPRIIEVRLLNPFLVKRMKRALGYLTDFPRILRRHNNKSFPA	191							
					:					
Db	97	LLLDNNFTGCLDLLDLALDSHPRIIEVRLLNPPSFRLLRPLGTYTDFSRLNRHNNKSFHY	156							
Oy	192	DNRATITLGSRNTGDEYFKVGEDPTVFADLLLATGSVVAGEVSHDEFDRYMASHSAHNATRII	251							

[illegible]

RESULT 6  
AE0636  
conserved hypothetical protein STY1185 [imported] - *Salmonella enterica* subsp. *enterica*  
C|Species: *Salmonella enterica* subsp. *enterica* serovar Typh  
A|Note: this species has also been called *Salmonella typhi*  
C|Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C|Accession: AE0636  
R|Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church  
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moun, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A|Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A|Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* se  
A|Reference number: AB0502; PMID:11677608  
A|Accession: AE0636  
A|Status: preliminary  
A|Molecule type: DNA  
A|Residues: 1-495 <PAR>  
A|Cross-references: GB:AL513382, PIDN:CAD08272.1; PID:g16502319; GSPDB:GN00176  
C|Genetics:  
A|Gene: STY1185

[illegible]











```

QY      351 TDALAKLVQDDIDVTVLNSLQATFVAHVHSGYVAKRREPLKAGIKYELQPHNAVPAIK 410
Db      359 LDSIKIALGVDVIMIPN--KPDHPFVWATLKNAASLDDAGYKVFHY----- 406
QY      411 DKGLTGSVSTSLHAKTFIVDKGRIFITGSFNDPPRSARLNTMGVYESPIAEOMERTLA 470
Db      407 DNGF-----LHSTVLIDDEILASVGTANNMDHRSFTLNFEVNAFTDQQLAKKLKQAFI 459
QY      471 D-----TSPEIAYR 479
Db      460 DDLAVSELTAKARYAKR 476

RESULT 17
H86774
cardiolipin synthase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: H86774
R:Boletun, A.; Wincker, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86774
A:Status: preliminary

```

## F90001

A:Cross-references: GB:AE005176; PID:q12733898; PIDN:AAK05058.1; GSPDB:GND0146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: cIsa  
C:Superfamily: Bacillus probable cardiolipin synthetase

Query Match	10.8%	Score 293.5;	DB 2;	Length 481;
Best Local Similarity	22.6%	Pred. No. 1e-13;		
Matches 110;	Conservative 83;	Mismatches 171;	Indels 123;	Gaps 17;

0y KVRRLDNTL-QIRHTPH-----TNGLSIDYLLNPNHEAFARAAMLESSEHSIDQY 104  
54 | :  
92 KMRQNHVIGLIHLMLVEEKSIISTN--TGKVFPGNRQKEFALIDDINKAHHNVHMEY 149

```
QY      105 YIMNDISGRLLFNLYVLAERGVAVRLLD---DNMTRGDLDLALLDSHPNIEVRLFN 161
        ||| | | : : : ||||| ||| | | : :
Db      150 YIFRNDRGHEIYDALLAAKRGVEVKLLDAMGSNTK-----MSNKELQ----- 196
QY      162 PFVLARKWALGYLDF-----PLNRBRHNSSTADNRATILIGRNIGDEYFVKE 212
```

Db 197. -----EAGGHAQVFFPLIMPLVNPRIINRYLRHRIIVIDIGTIIGYTGCGNVGDEVASITK 249

Oy 213 D-TVEADIDLIATGSGVGEVSHDPRKRWASHAHNATRIRSGNIGKGLOALGYNETSR 271  
 : | : | : | : | : | :  
Db 250 KFGYWRDHHLPLTDIVYSLQHRILDM-----NSQ 280

272 HALLRYETVEQSPLY-QKIDGRIDMOSVOTRLISDDPAKGLDRDRKRPPIAGRLDOL 330

```
Db      281 H---HHEVTEGEGYPDSIYEGH-----VATQLVTSGP-----DEDAEQIKLTYMKMI 325
```

QY 331 KQEKSVYLVSPFEVPTKSGIDALAKLVQDGIDIVLTNSLQATDVAAVHSGYKRRKPL 390Q

Db 326 SGAEREIIITPPYIPSDALHESIKLALLSGVQVKLLIPN--KPDHPLVYWTATYFHAADL 383

QY 391 LKAGIKYELQPNHAVPATKDKLTSSVTSLHAKTFIVDGKRIFGSEFNLDPRSRANT 450

Db 384 VKYGAQVY-----TYENG-----VHSKTLIDEPYASVGSANLDYRSLQJCF 426

451. EMGVVIESPKIAEDMERTLADTSPEYARVTLDRNRRLQWHDPATRKYYNEPEAKLMWR S10C

Db 427 EANWVLYDDYDISQLRN-----DENKDKLRSPLTLEREEERSKLVRFKE 471

QY 311 IAAHLS 31/  
| : : : :  
477 GIARITA 478  
Db





Db 394 RAGVYFRYPGF-----LHOKVVLVDRTDAVGSANLDRSRLNFE 436  
 QY 452 MGVIIESPKIAEQMERTLADTSPEYAVRYTLDRHN--RLQ 489  
 Db 437 VMVYVDEGFAGEVEMLEADFAESLEFTEPDRRSVRLQ 476

## RESULT 23

B84007

cardiolipin synthetase BH2858 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: B84007

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; M0ID:20512582; PMID:11058132

A:Accession: B84007

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-503 &lt;STO&gt;

A:Cross-references: GB:AP001516; GB:BA000004; MID:g10175192; PIDN:BA06577.1; GSPDB:GNOC

A:Experimental source: strain C-125

C:Genetics:

C:Superfamily: Bacillus probable cardiolipin synthetase

Query Match 10.1%; Score 274; DB 2; Length 503;  
 Best Local Similarity 24.3%; Pred. No. 2.9e-12;

Matches 109; Conservative 58; Mismatches 188; Indels 94; Gaps 12;

QY 61 ILDIRTPHTNGLS---DIYLLNDPHEAFARAALIESAHSLSLDQYIWRNDISGRLLF 117  
 Db 124 LFLAHLRLANPISFNTNTVLDGKETFAHIOALMAATHHILEYIYRDEIGCEIK 183  
 QY 118 NLVYLAERGVRLLDNDNTRGLDLLALDSHPNIEVRLPNFVLRKRALGYLTLDF 177  
 Db 184 EILMOKAKEGIHVRLVDGVGSMKLSYIQLDKAGVEIVPAPVYL----- 231  
 QY 178 P-----RLNRMHKSFPAADRATILGGRNIGDEYFKVGEOTVA---DLDLATGSVYGE 230  
 Db 232 PFINTHTYNNHKKIIVDTGVSGNLGDEX--LGKPPYFGFMDTLHYRGEAVRT 289  
 QY 231 VSHDFRYWASHAHNATRIIRSGNIGKGLALGYNDETSHALLRYRETEVQSPLOYKI 290  
 Db 290 LQILFLDMAHETG-----ETILKSTYLSPA 315  
 QY 291 QTRIDMQSVQTRLISDPKAGLDROKRPPIAGRLQDALKQPEKSYLVSPYFVPTKSG 350  
 Db 316 LTNMKDDGGVQ--MIASGP-----DTRWEINKKLFFSMITSAKKSIMTSPYIPDEDI 367  
 QY 351 TDALAKIVQGIQVTVTLNLSQATDVAAVHSGYVKRPPLKGIKRIYELQPNHNAVPAK 410  
 Db 368 LSAKLTAALSIDVRLVLPN--RPDKRIVHASSRSTYPLEAVKVEY----- 415  
 QY 411 DKGLTGSSTVSLAKTFIVDGKRIFGSNFLDPFSARLNTMGVIESPKIAEQMERTLA 470  
 Db 416 TRGF-----LHKKIIVNELASIGTSMNDMSFHLNEVNAF-----LYTKKS 459  
 QY 471 DTSEPAIVRYTLDRHNLQMHDPATKRTY 499  
 Db 460 VTLVSDVYVDLEHTNOIREQFRNRAMY 488

## RESULT 24

C84125

cardiolipin synthetase BH3803 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: C84125

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; M0ID:20512582; PMID:11058132

A:Accession: C84125

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 &lt;STO&gt;

A:Cross-references: GB:AP001520; GB:BA000004; MID:g10176401; PIDN:BA07522.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3803

Query Match 10.1%; Score 273.5; DB 2; Length 398;  
 Best Local Similarity 23.7%; Pred. No. 2.2e-12;

Matches 116; Conservative 66; Mismatches 191; Indels 117; Gaps 16;

QY 26 LILCLLSCSWLPLEERFERSRFRNFKSRVRLDNIQIHTPTNLSIYLNDPHEA 85  
 Db 5 VLFLVAILIWL--RPDYQGLKKQKGAVRNRQ--PIH-----GTVRLLPTEDEF 53  
 QY 86 FAARAALIESAHSLSLDQYIWRNDISGRLLFNLYLAERGVRLLD---DNNTRG 141  
 Db 54 IGSILHDKAEHHIHILEYIFRDHIGKKILQHEGKAKEGAVRLVDRFGADVNRS 113  
 QY 142 LDLLALDSHPNIEVRLPNFVLRKRALGYLTPRNLNRMHKSFPAADRATITGR 201  
 Db 114 IOSLKQA--GAQFEYAHRISSPY-----W-----FSLNRNRHKKITVIDKIGYIGY 160  
 QY 202 NIGDEYFKVGEPT---VFADLILATGSVYGEVSHDRVWASHASAMATRIIRSGNIGK 258  
 Db 161 NIGDEY--LGRDPKLGFWRYHILRTGDGVQDLODQFIOQWERSRLPVTR----- 209  
 QY 259 GLQALGYNDETSHALLRYRETEVQSPLOYKIQTRIDMQSVQTRLISDPKAGLDROK 318  
 Db 210 -----DRWLXPPLAKPHELR 226  
 QY 319 KPPIAGRLQDA---LKQPEKSYLVSPYFVPTKSGTDAKALQVQSDIVTVLNTSLQAT 374  
 Db 227 IPTNGSFLEDSFLQVQAEETITIGPYFIPGKELHALLDAARGAVALRVLP--KKG 284  
 QY 375 DVAAVHSGYVKRPPLKAGIKYLELOPNHNAVPAKTRKGLTGSSVTSLSAKTFVDDKRI 434  
 Db 285 DHPVKEAPAPYFELLEGGINITYRY-----RGF-----FHGAATYIDKLA 327  
 QY 435 FIGSFNLPFSARLNTMGVIESPKIAEQMERTLADTSPEYAVRYTLDRHNLQMHDP 494  
 Db 328 DVGIANFDKRSFRNTYINCLLYKEMIQVAREL-----DYDSIS-----RLQMDLA 378  
 QY 495 TRKTYPNEPE 504  
 Db 379 NRSEFHRTKE 388

## RESULT 25

S60089

cardiolipin synthetase homolog ywle - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 15-Feb-1996 #sequence\_revision 01-Mar-1996 #text\_change 20-Jun-2000

C:Accession: S60089; S55414; F70059

R:Crúz Ramos, H.; Bourcier, L.; Moszer, I.; Kunst, F.; Danchin, A.; Glaser, P.

EMBO J. 14, 5984-5994, 1995

A:Title: Anaerobic transcription activation in Bacillus subtilis: identification of

A:Reference number: S60080; M0ID:96112813; PMID:8846791

A:Accession: S60089

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-500 &lt;CRD&gt;

A:Cross-references: EMBL:249884; MID:g9711335; PIDN:CAA90049.1; PID:g9711345

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995

R:Glaser, P.; Danchin, A.

submitted to the EMBL Data Library, May 1995

A:Description: Cloning and sequencing of the Bacillus subtilis chromosomal region for

A:Reference number: S55414

A:Accession: S55414

A:Status: preliminary





QY 274 LARRETVESPLXOKIOTGRIDMOSVOTRLISDDPAKGLDRDRKRPPIAGRLQDALAKOP 333  
 Db 328 REGYVOPFADSP-----DDEPVGEIITMNLIS-----KA 357  
 QY 334 EKSVYLVSPFYVPTKSGTDALAKLVODGIDVTLNSLATDVAAYHSGVYKRPPLKA 393  
 Db 358 TKVYVITTPFLVIGNEMVATLTSAAKGVDVRIITPHI--PDKKIYHVSATYKSYKYLIES 415  
 QY 394 GIKLYELOPNHVAVPATKDKGLTSSVTSIAKTFIVDGKRIEFIGSFNDRSARLNTENG 453  
 Db 416 GVKIYEVMEGFI-----IHSKTYVCDNENGVGGSIMDRSRLYLHEGCG 458  
 QY 454 VYESKRIAEQOMERTLADTSPEYAVRVTLDNRHRLQMHDPATKTYTPNEPEALKRIRIA 513  
 Db 459 VMVYKATNYIYDIKKDFMDTL-DKSKETTELLEINKVWY-----STLM-RVVL 503  
 QY 514 KILSL 519  
 Db 504 RVFAPL 509

## RESULT 30

cardiolipin synthase [imported] - Nostoc sp. (strain PCC 7120)  
 A:Accession: AH2333

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AH2333  
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kultz, T.; Sasamoto, S.; Watanabe, A.; Iritaghi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Experimental source: strain PCC 7120

A:Residues: 1-480 <KUR>  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Cross-references: GB:BA000019; PIDN:BAH75922.1; PID:q17133358; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Gene: alr4223  
 C:Superfamily: Bacillus probable cardiolipin synthetase

Query Match 9.3%; Score 253; DB 2; Length 480;  
 Best Local Similarity 24.1%; Pred. No. 9, 6e-11;  
 Matches 111; Conservative 68; Mismatches 174; Indels 108; Gaps 16;

QY 68 PHTNGLSDIYLLNDPHEAFARALIESAHSIDLOYYIMRNDISGRLLFNLYLAERG 127  
 Db 115 PFTSG-NNAKLLJNGQGYAAMLSAASNSYLLQSYIVVDKAGNERKDALIAKAGQ 173  
 QY 128 VAVRLLDNNTRGLDLDLALDSHPNIEVRLFNPFVLRKWRALGYLTDPPRLNRHMK 187  
 Db 174 IRYLLLYDEIGSKISRLVYKSLQKDYIOVSFHTTRGRGNRP-----QINFNRHKK 225  
 QY 188 SFTADNRATILGGRNIGDEYFKVED---TFADILILANTGSVYGVSHDF--DRYWSH 242  
 Db 226 ILVYDORLAFIGLNSIDET--LGNKPLSPMDYTHMMLGPTVOSLQCCFLQDWAWAT- 282  
 QY 243 SAHNATRIIRSGNIGKGLQALGYNDETSRAHLRYRETVESPLYKIQIGRIDMOSVOT 302  
 Db 283 -----RQYIDVNNQVOP-----NMSDYT 301  
 QY 303 RLII-SDDPAKGLDRDRKRPPIAGRL--QDALOPEKSVYLVSPFYVPTKSGTDALAKVQ 359  
 Db 302 ALVFPPTGPA-----DKLK--ACKLEFVSAINQOQTRLIATIPYFVDDSTLTALAL 353  
 QY 360 DGDIVVTLNSIQATDVAAYHSGVYKRPPLKAGIKLYELOPNHVAVPATKDKGLTSSV 419  
 Db 354 RGVDAVRIIPN--RPDHLIVLVCSESYTEMKATNKLRYK--HGF----- 396  
 QY 420 TSLHAKTFIVDGKRIEFIGSFNDRSARLNTENG--MGVVISPKIAEQMERTLADTSPEYAV 478

Db 397 --MHQVILLIDKEMAGVGVNLDNRSEFLNEFVMEGVANSQFYKSVSEMLQADLKAALAV 454  
 QY 479 RVTLDRHNRNLQMHDPATKTYTPNEPEALKRIRIAKILSL 519  
 Db 455 -----DFSDYERKYLIMFKLAVIISLL 476

## RESULT 31

cardiolipin synthase VC1670 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: G82171  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.; Chaudron, D.; Ermolaeva, M.D.; Yamathayan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: AB2035; MUID:20406833; PMID:10952301

A:Accession: G82171  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-484 <HEI>

A:Cross-references: GB:AE004245; GB:AE003852; NID:9656183; PIDN:AAF4820.1; GSPDB:GN  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Gene: VC1670  
 A:Map position: 1  
 C:Superfamily: Bacillus probable cardiolipin synthetase

Query Match 9.2%; Score 248.5; DB 2; Length 484;  
 Best Local Similarity 24.6%; Pred. No. 2, 1e-10;  
 Matches 110; Conservative 64; Mismatches 160; Indels 133; Gaps 20;

QY 57 RLDNLIQIR-HPHTNGLSDIYLLNDPHEAFARALIESAHSIDLOYYIMRNDISGRLL 115  
 Db 102 RIDELCNMNGMIPALSG-NILSLNPNELIHAIIDIERAQPIQIMVYIHHGSLADA 160  
 QY 116 LFNLYLAERGVRRLLDNDNT-----RGLDLALDSHPNIEVRLFN 161  
 Db 161 VASAVIQASKRGVNVKLLDSAGSPFRFSPEMKWRDAGI--EVQALREVSP----- 211  
 QY 162 PFVLRKWRALGYLTDPPRLNRHMKSFADNRATILGGRNIGD-EYK-----VEGDYVF 216  
 Db 212 -----WRIF-----LRLDLRQHRKIYIDETIATGSMMDVAFQONAGVGO--W 257  
 QY 217 ADDILATGSVGEVS--HDFRYWASHAHNATRIIRSGNIGKGLQALGYNDETSRAHL 274  
 Db 258 IDIMVAVTGPVNVLSAHCWD--WEET-----GSRMLPKNDE----- 294  
 QY 275 LARRETVESPLXOKIOTGRIDMOSVOTRLISDDPAKGLDRDRKRPPIAGRLQDALAKOPE 334  
 Db 295 CRLEPNQOPHP1-QVVPISG--PGMBENLISQ-----VTLAIQAN 332  
 QY 335 KSVYLVSPFYVPTKSGTDALAKLVODGIDVTLNSLATDVAAYHSGVYKRPPLKAG 394  
 Db 333 KSVCTTPFYVBSADLLATLAKMTAQRGKIKVDIIP--KNDLSLAWQASRAATYGGELLEAG 390  
 QY 395 IIKLYELOPNHVAVPATKDKGLTSSVTSIAKTFIVDGKRIEFIGSFNDRSARLNTENG 454  
 Db 391 VQIHFF-----DGL-----LHTKSVIIDQRCCLVGYVNLDMRSILMLNFELTL 433  
 QY 455 VIESPKIAEQMERTLADTSPEYAV 478  
 Db 434 AVDDLEFQOQHMLDOOYIDOSHVVY 460

## RESULT 32

cardiolipin synthetase (EC 2.7.8.-) [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC0266



Query Match	8.8%	Score 239;	DB 2;	Length 486;
Best Local Similarity	22.0%	Pred. No. 1e-09;		
Matches 121; Conservative	77;	Mismatches 189;	Indels 162;	Gaps 23

RESULT 33  
F86771  
cardiolipin synthase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

A;Status: preliminary

Query Match	8.8%	Score 238.5;	DB 2,	Length 487;
Best Local Similarity	22.3%	Pred. No. 1.1e-09;		
Matches 99;	Conservative 62;	Mismatches 169;	Indels 113;	Gaps 14;

RESULT 34  
CEE410

```
C:\Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
```

A;Status: preliminary

A;Residues: 1-398 &lt;GLA&gt;

A; Cross-references: EMBL:Z49782; NID:9853752; PIDN:CA98966.1; PID:9853756566.1; R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, J.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carr, A.; Ehrlich, S.D.; Emmertson, P.T.; Entian, K.D.; Errington, J.; Fábreg, C.; Nature 390, 249-256, 1997

iech, J. J., Harwood, C. R., Henauf, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinolet, A., A:Authors: Lauber, J., Lazarevic, V., Lee, S.M., Levine, A., Liu, H., Masuda, S., Maun, R., Y. M., Ogawa, K., Ogizawa, A., Ouegga, B., Park, S.H., Parro, V., Pohl, T.M., Poterstein, R., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, V., Sato, T., Scatena, A:Authors: Schleich, S., Schriener, R., Scoffone, F., Sekiguchi, J., Sekowska, A., A:Authors: kench, M., Tamashiro, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, T., Winters, P., Wipac, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: C70060  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-398 <KUN>  
 A:Cross-references: GB:299123; GB:AL009126; NID:92636240; PIDN:CAB15747.1; PID:92636256  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ywje  
 C:Superfamily: Bacillus probable cardiolipin synthetase

Query Match 8.5%; Score 231; DB 2; Length 398;  
 Best Local Similarity 21.6%; Pred. No. 3e-09;  
 Matches 94; Conservative 65; Mismatches 169; Indels 108; Gaps 13;

OY 74 SDIYLLNDPHEFAARAALIESAHSIDLQYIMRNDISGRLLFNLYLAERGVRL 133  
 DB 43 SDIELHGCADLVERMMDIROAASSVHMFIMKNDVSHMWTLLKTKAAGVSYLL 102  
 OY 134 LDDNNTRGIDLLALDHPNIEVRLFN---PVLKRRKALGYLTDFPRLNRRMHNKSF 189  
 DB 103 LDMACGRALIKTALDTMKAGVHVHNNRPREPF-----FFHMKRRHRTIT 150  
 OY 130 TADNATILGKNGIDEPFKVGEDVFADLILATGVSVEVSHDFRYWASHAHNATR 249  
 DB 151 VLDGIGYIGGFENIAEEY--LGKKAKFGWMEYHL--RMIGEGVHDLQTLFASDLKRT-- 205  
 OY 250 IIRSNIGKGLDLYNDTSRHALRREYEQSPLOKIQDTGR-----DMQSVOT 302  
 DB 206 -----GLE-LG-----SDWPKLQGGTSHKIVATDGLSEN 236  
 OY 303 RLISDDPAKGLDRDRKPRPIAGRLDALKQPEKSVYLVSPVEPTKSGTDALAKLVODGI 362  
 DB 237 IYLAN-----IAQAKNRLTVCPPYIPSPQDEALINARKNGV 274  
 OY 363 DVTVLNTSLQATDVAHVSHGYKRYKKPLKAGIKLEYLQPNHNAVATKDKGLTGSSVSL 422  
 DB 275 SVRIIVP--MMSDHPVREAAFTYVSELLDACLIYRYQGF-----Y 315  
 OY 423 HAKTFIVGKRIFFISFNLDPRSARLNTMGVIES-----PKIAQOMERTLADTSP 474  
 DB 316 HVKALIIDHLSIIGTANFDRKSLFLNEVYVEIDDEAFTSEVYATIEEDMKKSELLTWE 375  
 OY 475 EYAVRYTLDRHNRQW 490  
 DB 376 DESKRTF--RQRPAM 389

## RESULT 35

83377

probable phospholipase PA2155 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*

C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: B83377

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10964043

A:Accession: B83377

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-401 <STO>

A:Cross-references: GB:AE004642; GB:AE004091; NID:99948163; PIDN:AA05543.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2155

Query Match 8.5%; Score 231; DB 2; Length 401;  
 Best Local Similarity 23.5%; Pred. No. 3e-09;

Matches 100; Conservative 58; Mismatches 178; Indels 90; Gaps 12;  
 OY 76 IYLLNDPHEFAARAALIESAHSIDLQYIMRNDISGRLLFNLYLAERGVRL 135  
 DB 11 VELLNGEYFFRLRQCIARREILLETFTFEDEVGQLOEALSAAREGEVQVTD 70  
 OY 136 DNNTRGIDLLALDHPNIEVRLFNPEVLRKRRALGYLTDFPRLNRRMHNKSF 195  
 DB 71 GVTASLSPDYLARLASCVRVHLFDP-----RPLRGMTN---LFRRLHRLVYIDRRQ 123  
 OY 196 TILGGRNIGDEYFKVGEDVFADLILATGVSVEVSHDFRYWASHAHNATRIISGN 255  
 DB 124 AFGGINTGEDH-----LVRRGN 141  
 OY 256 ICK---GICALYNDTSRHALRREYEQSPLOKIQDTGRIDMQSVOTRLISDDPAKG 312  
 DB 142 MAKQDYAVRVEGPPVVDIROACLALLEPRADYPLRPSAG---QPARVRLV----- 191  
 OY 313 LDRDRRRPPIAGRLDALKQPEKSVYLVSPVEPTKSGTDALAKLVODGIDVTYLSLQ 372  
 DB 192 RNDQSSDIEEYVLOAIROARRRLINAYFPFGYRLRLRELRDAARGVRVLYQGM 251  
 OY 373 ATDVAHVSHGY--KYRKPLKAGITLYEL--QPNNAVATKDKGLTGSSVSLHAKTFIV 429  
 DB 252 DMPVLRCLSRLLYDY--LLREGVRIHEQCRP-----LHGKVAVI 289  
 OY 430 DCKRIFFISFNLDPRSARLNTMGVIESPKIAEQMERTLADTSPAYRVTLDRHNR-- 487  
 DB 290 DDDWSTPSSNLDPLSLNLNLANVINDVANQGLYHRLRLARHCRIS--RRHARRG 348  
 OY 488 LQWHP 493  
 DB 349 YWWRAP 354

## RESULT 36

A10598

probable phospholipase ybho [imported] - *Salmonella enterica* subsp. *enterica* serovar

C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh

A:Note: this species has also been called *Salmonella typhi*

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001

C:Accession: A10598

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair

, S.; Moulé, S.; O'Gaora, P.

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Reference number: AB0502; PMID:11677608

A:Accession: A10598

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-413 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05259.1; PID:916502026; GSPDB:GN00176

C:Genetics:

A:Gene: ybho

Query Match 8.4%; Score 229; DB 2; Length 413;  
 Best Local Similarity 22.2%; Pred. No. 4.4e-09;  
 Matches 92; Conservative 55; Mismatches 163; Indels 104; Gaps 11;

OY 74 SDIYLLNDPHEFAARAALIESAHSIDLQYIMRNDISGRLLFNLYLAERGVRL 133  
 DB 9 NQIQLLENGDQFYPAVFAIAQAOQKILLETFTFEDEVGKRIHALKAQGVKAEVL 68  
 OY 134 LDDNNTRGIDLLALDHPNIEVRLFNPEVLRKRRALGYLTDFPRLNRRMHNKSF 193  
 DB 69 LDGYSPSLSDAFVGLTSAGVIFRYTP---RPLGLGRTN---LFRMRHRIIVIDD 121  
 OY 194 RATILGGRNIGDEYFKVGEDVFADLILATGVSVEVSHDF-----RYWASH 242  
 DB 122 RIAFGGINYSHEMSDYGPOAKQDYAVRVEGPPVADIL--QFEVNLPGQSPARRMWRKH 180

cardiolipin synthase [imported] - Agrobacterium tumefaciens. (strain C58, Dupont)

A:Gene: C1S; 00366 .  
A:Genetic code: SGC3

GSIDB:GN

```

OY      245  HNAIRIRSGNIGKGLALGYNDETSRHALLRYRETVEOSPLYKQIQTGRIDWQSVTRL 308
          | | : | | |
Db      181  HKAEENRQPG-----EAOVLLVMR----- 199

```

```
QY      305  ISDDPAKGLDRDRKPP IAGRLQDALOKPEKSVYLVSPYFVPTKSGTDALAKLVODGIDV 366
          | : | | : : | | | : : | :
Db      200  -----DNEEHRRDIERHYLKMQLQARREVIITANATFFPGYRFLHALKRAARRGRI 250
```

OY 365 TVLTNSLQATDVAALVHSGYKXRRPLLKAGIKLYELOPNHAAVPATKDKGLTGSSVTSLHA 424  
:: | : | : | : | : | :  
Db 251 KLIIQG--EPDMPIRVGARGALLNYLVKGGVOVFEEYRRR-----PLHG 291

Qy	425	KTFIVDGKRIFFIGSNLDPRSARLNTEMGVIESPKIAEQMERTLAD	471
	:::	:	:
Db	292	KVALMDHWAATVGSSNLDPLSLNLNEANVIITHD---	RHFNQTLRD 334

Search completed: May 12, 2003, 10:00:24  
Job time : 53 secs

18

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:55:17 ; Search time 91 seconds  
(without alignments)  
1188.733 Million cell updates/sec

Title: US-10-066-551-4  
Perfect score: 2713  
Sequence: 1 MRANKTQAMPSETISLMKT.....KLMKRIAKILSLPIESLL 525

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_organella:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_proteic:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2713	100.0	525	09JTT1	09JTT1 neisseria m
2	2599	95.8	508	09JYU0	09JYU0 neisseria m
3	1113.5	41.0	493	08X917	08X917 escherichia
4	1106.5	40.8	514	09CKM5	09CKM5 pasteurella
5	1100	40.5	494	08Z028	08Z028 salmonella
6	1100	40.5	495	08Z7M0	08Z7M0 salmonella
7	824.5	30.4	466	09BNZ3	09BNZ3 rhizobium l
8	821	30.3	518	08UEX3	08UEX3 agrobacteri
9	816.5	30.1	466	0984B3	0984B3 rhizobium l
10	808.5	29.8	529	09HTP4	09HTP4 pseudomonas
11	761	28.1	521	09Z0Q3	09Z0Q3 rhizobium m
12	427	15.7	652	09PEE1	09PEE1 xyloella fas
13	318	11.7	504	09ZPF5	09ZPF5 listeria in
14	311	11.5	504	08YAV5	08YAV5 listeria mo
15	308.5	11.4	494	09SGS9	09SGS9 staphylococ
16	293.5	10.8	481	09CGY2	09CGY2 lactococcus

17	293.5	10.8	482	16	09Z7Z0	09Z7Z0 listeria in
18	292.5	10.8	480	16	08XP94	08XP94 clostridium
19	289.5	10.7	485	16	08XMD5	08XMD5 clostridium
20	288.5	10.6	482	16	08YAE3	08YAE3 listeria mo
21	286.5	10.6	489	16	08XQ11	08XQ11 raietonia s
22	283	10.4	479	16	08R663	08R663 fuscobacteri
23	281.5	10.4	481	2	08RK18	08RK18 pseudomonas
24	275	10.1	505	16	09JUF8	09JUF8 staphylococ
25	274.5	10.1	480	16	09HFM0	09HFM0 pseudomonas
26	273.5	10.1	388	16	09K6C6	09K6C6 bacillus ha
27	264.5	9.7	492	16	08YDD8	08YDD8 brucella me
28	262.5	9.7	467	16	09PE19	09PE19 xyloella fas
29	261	9.6	359	16	09HWM2	09HWM2 pseudomonas
30	258	9.5	510	16	097E04	097E04 clostridium
31	253	9.3	466	2	09FHK4	09FHK4 zymomonas m
32	253	9.3	480	16	08YFHI	08YFHI anabaena sp
33	251	9.3	397	2	09LAE8	09LAE8 bacillus ce
34	250.5	9.2	422	16	08Y266	08Y266 raietonia s
35	248.5	9.2	484	16	09KRM2	09KRM2 vibrio chol
36	242.5	8.9	487	16	09BHI8	09BHI8 rhizobium l
37	241	8.9	395	16	09K251	09K251 streptomyce
38	239	8.8	486	16	08ZEL2	08ZEL2 yersinia pe
39	238.5	8.8	487	16	09CGC3	09CGC3 lactococcus
40	238	8.8	511	16	08XNG8	08XNG8 clostridium
41	231	8.5	401	16	0911W0	0911W0 pseudomonas
42	231	8.5	413	16	08ZOP4	08ZOP4 salmonella
43	229	8.4	413	16	08Z883	08Z883 salmonella
44	227	8.4	487	16	09Z034	09Z034 rhizobium m
45	227	8.4	492	16	08OCJ8	08OCJ8 agrobacteri

## ALIGNMENTS

RESULT 1	PRELIMINARY:	PRT:	525 AA.
ID 09JTT1			
AC 09JTT1			
DT 01-OCT-2000 (TREMBLrel. 15, Created)			
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE Phospholipase D-family protein			
GN NMA1646			
OS Neisseria meningitidis (serogroup A).			
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX NCBI_TaxID=65699;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;			
RX MEDLINE=20222556; Pubmed=10761919;			
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S., Jagsels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;			
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491".			
RL Nature 404:502-506(2000).			
DR EMBL AL662756; CAB6474-1;			
DR InterPro: IPR001736; PDB			
DR Pfam: PF00614; PDC: 2.			
DR SMART: SM00155; PDC: 2.			
KW Complete proteome.			
SQ SEQUENCE 525 AA; 59305 MW; 1CFB7AC5B82F1B02 CRC64;			
Query Match	100.0%;	Score 2713;	DB 16; Length 525;
Best Local Similarity	100.0%;	Pred. No. 4.4e-198;	
Matches 525;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
OY 1 MRANKTQAMPSETISLMKTRSLISLCLLCSSSWLPLEERTESRHNTSKPVRLDN 60			
DB 1 MRANKTQAMPSETISLMKTRSLISLCLLCSSSWLPLEERTESRHNTSKPVRLDN 60			

QY 61 IIQIRTPHTNGISLDIYLLNDPHEAFARALIESAHSLSLDQYITWRNDISGRLLFNLY 120  
 DB 61 IIQIRTPHTNGISLDIYLLNDPHEAFARALIESAHSLSLDQYITWRNDISGRLLFNLY 120  
 QY 121 YLAERGVAVRLLDNDNRGLDLDLLADSHPNIEVRLFNPEVLRKRWALGYLDNDFRL 180  
 DB 121 YLAERGVAVRLLDNDNRGLDLDLLADSHPNIEVRLFNPEVLRKRWALGYLDNDFRL 180  
 QY 181 NRRMHNKSTADNRATILGGRNIGDEYFVGEDEVFADLLATGSGVGEVSHDFDRYA 240  
 DB 181 NRRMHNKSTADNRATILGGRNIGDEYFVGEDEVFADLLATGSGVGEVSHDFDRYA 240  
 QY 241 SSSAHNATRIISGNIGKLGALGYNDERSHALLRYRETVEOSPLYOKIOTGRIDMOSV 300  
 DB 241 SSSAHNATRIISGNIGKLGALGYNDERSHALLRYRETVEOSPLYOKIOTGRIDMOSV 300  
 QY 301 QTRLLSDPAKGLDRRKPPIAGRLQDALKQPEKSVLYSPYFVPTKSGTDALAKLYOD 360  
 DB 301 QTRLLSDPAKGLDRRKPPIAGRLQDALKQPEKSVLYSPYFVPTKSGTDALAKLYOD 360  
 QY 361 GIDVYVLTNSLQATDVAHVSQVYKRPDLNAGIKLYELOPNHAPATKDKGLTGSVYT 420  
 DB 361 GIDVYVLTNSLQATDVAHVSQVYKRPDLNAGIKLYELOPNHAPATKDKGLTGSVYT 420  
 QY 421 SLHAKTFIYDGRKRIFGSNLDPKRSARLNTEMGVIESPKIAEQMERTLADTSPAYAYV 480  
 DB 421 SLHAKTFIYDGRKRIFGSNLDPKRSARLNTEMGVIESPKIAEQMERTLADTSPAYAYV 480  
 QY 481 TLDRHRLQMDPATRKTYPNPEAKLMKRIAKILSLPIESLL 525  
 DB 481 TLDRHRLQMDPATRKTYPNPEAKLMKRIAKILSLPIESLL 525

## RESULT 2

QY 09JYU0 PRELIMINARY: PRT: 508 AA.  
 ID 09JYU0: 09JYU0:  
 AC 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cardiolipin synthetase family protein.  
 GN NMB1434.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MC58 / SEROGROUP B:  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelein H., Saunders N.-J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Nelson J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scariello V., Masignani V., Pizzia M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappunli R., Venner J.C.,  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AE002493; AAF41795.1; -.  
 DR TIGR: NMB1434; -.  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 508 AA; 57350 MW; A2DE6CAC47CA25D5 CRC64;

Query Match 95.8%; Score 2599; DB 16; Length 508;  
 Best Local Similarity 98.8%; Pred. No. 2e-189;  
 Matches 502; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 18 MKTRSLISLLCLLSCSSWMLPPEERTESRHFNFSKPYRLDNLIIQIRTPHTNGISLDIY 77  
 DB 1 MKTRSLISLLCLLSCSSWMLPPEERTESRHFNFSKPYRLDNLIIQIRTPHTNGISLDIY 60  
 QY 78 LNDPHEAFARALIESAHSLSLDQYITWRNDISGRLLFNLYLAERGVAVRLLDND 137  
 DB 61 LNDPHEAFARALIESAHSLSLDQYITWRNDISGRLLFNLYLAERGVAVRLLDND 120  
 QY 138 NTRGLDLDLLADSHPNIEVRLFNPEVLRKRWALGYLDNDFRLNRRMHNKSTADNRATI 197  
 DB 121 NTRGLDLDLLADSHPNIEVRLFNPEVLRKRWALGYLDNDFRLNRRMHNKSTADNRATI 180  
 QY 198 LGRNIGDEYFVGEDEVFADLLATGSGVGEVSHDFDRYAASSAHNATRIISGNIG 257  
 DB 181 LGRNIGDEYFVGEDEVFADLLATGSGVGEVSHDFDRYAASSAHNATRIISGNIG 240  
 QY 258 KGLQALGYNDERSHALLRYRETVEOSPLYOKIOTGRIDMOSVQTRLLSDPAKGLDRDR 317  
 DB 241 KGLQALGYNDERSHALLRYRETVEOSPLYOKIOTGRIDMOSVQTRLLSDPAKGLDRDR 300  
 QY 318 RRPPIAGRLQDALKQPEKSVLYSPYFVPTKSGTDALAKLYODGIDVYVLTNSLQATDVA 377  
 DB 301 RRPPIAGRLQDALKQPEKSVLYSPYFVPTKSGTDALAKLYODGIDVYVLTNSLQATDVA 360  
 QY 378 AVHSGVYKRPDLNAGIKLYELOPNHAPATKDKGLTGSSTSLHAKTFIYDGRKRIFG 437  
 DB 361 AVHSGVYKRPDLNAGIKLYELOPNHAPATKDKGLTGSSTSLHAKTFIYDGRKRIFG 420  
 QY 438 SFNLDPRSARLNTEMGVIESPKIAEQMERTLADTSPAYAYVTLDRHRLQMDPATRK 497  
 DB 421 SFNLDPRSARLNTEMGVIESPKIAEQMERTLADTTPAYAYVTIDRHRNLOMDPATRK 480  
 QY 498 TYNPEPEAKLMKRIAKILSLPIESLL 525  
 DB 481 TYNPEPEAKLMKRIAKILSLPIESLL 508

## RESULT 3

QY 08X917 PRELIMINARY: PRT: 493 AA.  
 ID 08X917: 08X917:  
 AC 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putative synthase.  
 GN YMDC OR 21680 OR ECS1424.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.,  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 DR [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsuko E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.,  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).

DR EMBL: AE005315; AAG55792.1; -  
 DR EMBL: AP002555; BAB34847.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 DR Complete proteome.  
 KW SEQUENCE 493 AA; 55929 MW; DECA07F83D7D1CFF CRC64;

Query Match 41.0%; Score 1113.5; DB 16; Length 493;  
 Best Local Similarity 48.1%; Pred. No. 2.6e-76;  
 Matches 223; Conservative 82; Mismatches 142; Indels 17; Gaps 5;

Qy 72 GLSDIYLLNDPHEAPARAALIESAHSIDLOYYIMRNDISGRLLFNLYLAERGVR 131  
 Db 37 GCGGFLPEKSLDAFAARYRLAEMSEHLDVQYIIMQDMSRLLEFALLAAKRGVR 96  
 Qy 132 LLDDNNTRGLDLDLLADSHRNIEVRLFPVLRKRALGYLTDFPRLRNHNKSFPA 191  
 Db 97 LLDDNNTRPGLDLIRLDSHPRIEVRLEFPESFLRLPLGYITDFSRLRNHNKSFV 156  
 Qy 192 DRRATILGGRNIGDEYFKEVDYFADLDILATGSVYGEVSHDFEDRYASHANATRII 251  
 Db 157 DGVTILVGRNIGDAYFAGGEPLFSDLVMAIGPVEVDADDFARYWCKSVPLQOVL 216  
 Qy 232 RSGNIGKG-----LQALGYNDETSRHALLRYRETVEOSPPLYOKIOTGRIDMOVSOTRL 304  
 Db 217 ---DVPEGMADRIELPASMNDAMTH---RYLRKMESPRLNHLVDGTLPLIMAKTRL 269  
 Qy 305 ISDDPAKGLDRRRKRPPLAGRLQDALKOPEKSVYLVSPFVPTKSGTDALAKLVODGITY 364  
 Db 270 LSDDDAKGKGRKRRSLPQRLFDLIMGSPSERIDIIISYFVPTRAVQALLRVRKGVK 329  
 Qy 365 TVLTNSLQATDVAAYHSGVYKRPPLKAGIKLYELOPNHAAVPAAT-KDKGLTGSVTSLSH 423  
 Db 330 ALTNSTLANDVAAYHAGARRRKKLLRGVELYELKPTREOSSTLHDGTCITONSASLSH 389  
 Qy 424 AKTFIVDGKRIFIGSFNDPRSARLNTENGVAIVESPKIAEOMERTLADTSPEYAVRVTL 483  
 Db 390 AKTFEIDGKTVFIFGSFNEPRSTLLNTEMGFIESETLQALDKRPIQOQYDAAMQLRID 449  
 Qy 484 RHNRLQWHDPRTRK--TYNEPEAKLMKRIAKTILSLPIESLL 525  
 Db 450 RMRGRLNWDRRAKKEIVLKEPATSFWRKRVNLASILPEVWLL 493

## RESULT 4

O9CKM5 PRELIMINARY; PRT; 514 AA.

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical protein PM1586.  
 GN PM1586.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Pasteurella.  
 OX NCBI\_TaxID=747;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PM70;  
 RA MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida PM70."  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
 DR EMBL: AE006196; AAK03670.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 514 AA; 57882 MW; C65B0EB65D80F216 CRC64;

Query Match 40.8%; Score 1106.5; DB 16; Length 514;

Best Local Similarity 46.4%; Pred. No. 9.4e-76;  
 Matches 238; Conservative 80; Mismatches 170; Indels 25; Gaps 8;

Qy 23 LISLCLLLCSCSSWLPLEERKESRHNFSKVRLDNLI-----QIRTPPTNGISDI 76  
 Db 17 LFSLVITALISYOR-LP-----TSDRPISTHNPPLNGLRLRILRQVQHPPLTGLG--- 67  
 Qy 77 YLLNDPHEAPARAALIESAHSIDLOYYIMRNDISGRLLFNLYLAERGVRRLDLD 136  
 Db 68 YPLGDKDAFLARLALSEHAHENTLDQYYIMHNDVSGHLLQSLYKAAARGVAVRLDLD 127  
 Qy 137 NNTRGDLDLLADSHRNIEVRLFPVLRKRALGYLTDFPRLRNHNKSFPAADNRAT 196  
 Db 128 NNTKGMDTILASLNAPNIQIRLFNPFMQRQYRWLGFSLDFEFLNRNHNKSFPAAGVMS 187  
 Qy 197 ILGGRNIGDEYFKEVDYFADLDILATGSVYGEVSHDFEDRYASHANATRIISGNT 256  
 Db 188 ILGGRNIGDEYFKEVDYFADLDVATGAVVNIQITDFDRYNSPSSYLESIIRDP 247  
 Qy 257 GKGLQALGYNDETSRHALLRYRETVEOSPPLYOKIOTGRIDMOVSOTRLISDDPAKGLDR 316  
 Db 248 TRNPRLPADDETQ-----TYLKQLELPRAKSLKAGTLAFTWAEELLISDDPKALGKS 302  
 Qy 317 RRRPRLAGRLQDALKOPEKSVYLVSPFVPTKSGTDALAKLVODGIDVYLTNSLQATDV 376  
 Db 303 LIQDSVLAHLAPYMLNAKNLILVSPYFPTHVGVDFLSISQGTQVSLTNSLEATDV 362  
 Qy 377 AAVHSGVYKRPPLKAGIKLYELOPNHAAVPAATKDG--LTGSSVSLAKTIVDGKRI 434  
 Db 363 STVHSGVYKRPPLKAGIKLYELOPNHAAVPAATKDG--LTGSSVSLAKTIVDGKRI 421  
 Qy 435 FIFSGFLDPRSARLNTENGVAIVESPKIAEOMERTLADTSPEYAVRVTLDRHNRLQWHDPA 494  
 Db 422 FVSGFMDRSMALNTEMGILLDSPELRLISGLDQONQANVAFSVKLNDAQALYMETQE 481  
 Qy 495 TRK--TYNEPEAKLMKRIAKTILSLPIESLL 525  
 Db 482 NGKRIYENEPHTSWPKRFESVWCLSWLPVEHLL 514

## RESULT 5

O8ZQ28 PRELIMINARY; PRT; 494 AA.

DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative phospholipase.  
 GN YMDC OR STM1148.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McGlelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2."  
 RT Nature 413:852-856(2001).  
 DR EMBL: AE008749; AAL20078.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 494 AA; 56217 MW; 571C01EFC4FCB288 CRC64;

Query Match 40.5%; Score 1100; DB 16; Length 494;

Best Local Similarity 46.9%, Pred. No. 2.7e-75;  
Matches 227; Conservative 80; Mismatches 149; Indels 28; Gaps 7;

QY 57 RLDNLTQ---IRHPHTGLSDIYLLNDPHEAFARAALIESAHSIDLQYIWNDSIG 113  
D 24 RLRAVEPLCARHP---GEGCILLDNSLDFAFARYRLTEMAATLDVQYIWNEDMSG 79  
QY 114 RLLENLYLAERGVYRLLDDNNTGLDILLALDSHPNIEVRLFPVLRKRALGY 173  
D 80 RLLESVLLSAARKGVYRLLDDNNTGLDILLALDSHPNIEVRLFPVLRKRALGY 139  
QY 174 LTDFRLNRMRHNSFTADNRATILGGRNIGDEYKVEDYFADLDILATGVSVEVSH 233  
D 140 LTDFRLNRMRHNSFTADNRATILGGRNIGDEYKVEDYFADLDILATGVSVEVSH 199  
QY 234 DFDRYWASHAHNATRIIRSGNIGKGLQAL-----GYNDETSRIALLRYRETVESQ 284  
D 200 DFERWRCSSVSTLQOVLSE-----OELTORLELPESWYNDETR---RYLHKLETS 250  
QY 285 PLYOKIOTGRIDMOSVOTRLISDDPAKGLDRRRKPIAGRLQDALKQPEKSVLYSPYF 344  
D 251 QPMADLRGTPLWAKTRLLSDPSKGEKQAQRHSLLPQRLFDVWGSPTERIDITSAVF 310  
QY 345 VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 404  
D 311 VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 370  
QY 405 A-VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 463  
D 371 EHERAVHNRGLTNGSSGLHAKTFISDGSKYFISGLNDPSTLNTMGVYIESETLAT 430  
QY 464 QMERLADTSEYAVRYVLDNRNLQWHD--PATRYTPNPEPAKLMKRIAKIISLPI 521  
D 431 LIHKFTQSORDAAMQRLDRGRINWIDROOEKVKLKEPATRFWRVYRLAAILPV 490  
QY 522 ESL 525  
D 491 EWL 494

RESULT 6  
ID 0827N0 PRELIMINARY: PRT: 495 AA.  
AC 0827N0;  
DT 01-MAR-2002 (Tremblrel, 20, Created)  
DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)  
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)  
DE Hypothetical protein str1185.  
GN str1185.  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
NCBI\_Taxid=601;  
RP SEQUENCE FROM N.A.  
RC MEDLINE-21534947; PubMed-11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahina M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Felwell T., Hanlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,  
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,  
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrrell B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18.";  
RL Nature 413:848-852(2001).  
DR EMBL: AL627269; CAD08272.1; -  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC; 2.  
DR SMART: SM00155; PLDC; 2.  
KW Hypothetical protein; Complete proteome.

SO SEQUENCE 495 AA; 56348 MW; 45552578DB83AB9 CRC64;

Query Match 40.5%, Score 1100; DB 16; Length 495;  
Best Local Similarity 46.9%, Pred. No. 2.8e-75;  
Matches 227; Conservative 80; Mismatches 149; Indels 28; Gaps 7;

QY 57 RLDNLTQ---IRHPHTGLSDIYLLNDPHEAFARAALIESAHSIDLQYIWNDSIG 113  
D 24 RLRAVEPLCARHP---GEGCILLDNSLDFAFARYRLTEMAATLDVQYIWNEDMSG 80  
QY 114 RLLENLYLAERGVYRLLDDNNTGLDILLALDSHPNIEVRLFPVLRKRALGY 173  
D 80 RLLESVLLSAARKGVYRLLDDNNTGLDILLALDSHPNIEVRLFPVLRKRALGY 140  
QY 174 LTDFRLNRMRHNSFTADNRATILGGRNIGDEYKVEDYFADLDILATGVSVEVSH 233  
D 141 LTDFRLNRMRHNSFTADNRATILGGRNIGDEYKVEDYFADLDILATGVSVEVSH 200  
QY 234 DFDRYWASHAHNATRIIRSGNIGKGLQAL-----GYNDETSRIALLRYRETVESQ 284  
D 201 DFERWRCSSVSTLQOVLSE-----OELTORLELPESWYNDETR---RYLHKLETS 251  
QY 285 PLYOKIOTGRIDMOSVOTRLISDDPAKGLDRRRKPIAGRLQDALKQPEKSVLYSPYF 344  
D 252 QPMADLRGTPLWAKTRLLSDPSKGEKQAQRHSLLPQRLFDVWGSPTERIDITSAVF 311  
QY 345 VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 404  
D 312 VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 371  
QY 405 A-VPTKSGTALAKLYODGIDVYLTNSLOATDYAAVHSGVYRKRLKAGIKLYELOPNH 463  
D 372 EHERAVHNRGLTNGSSGLHAKTFISDGSKYFISGLNDPSTLNTMGVYIESETLAT 431  
QY 464 QMERLADTSEYAVRYVLDNRNLQWHD--PATRYTPNPEPAKLMKRIAKIISLPI 521  
D 432 LIHKFTQSORDAAMQRLDRGRINWIDROOEKVKLKEPATRFWRVYRLAAILPV 491  
QY 522 ESL 525  
D 492 EWL 495

RESULT 7  
ID 098NZ3 PRELIMINARY: PRT: 466 AA.  
AC 098NZ3;  
DT 01-OCT-2001 (Tremblrel, 18, Created)  
DT 01-OCT-2001 (Tremblrel, 18, Last sequence update)  
DT 01-OCT-2001 (Tremblrel, 18, Last annotation update)  
DE Phospholipase D-family protein.  
GN MLR9675.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
NCBI\_Taxid=381;  
RP SEQUENCE FROM N.A.  
RC MEDLINE-21082930; PubMed-11214968;  
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.";  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP003017; BAB54862.1; -  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC; 2.  
DR SMART: SM00155; PLDC; 2.









QY 492 DPATKTYNPEAKIMK--RIAAKLSLLPIESTL 524  
DB 561 --APRRKLPG-----LYKLNYSVGKLSKALPYLDL 588

RESULT 13  
Q92F05 PRELIMINARY; PRT; 504 AA.  
AC 092F05;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypothetical protein lin0008.  
GN LIND008.  
OS Listeria innocua.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1642;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-CLIP 11262 / SEROVAR 6A;  
RX PubMed=11679669;  
RA Baguero F., Berche P., Bloecher C., Rusniok C., Amend A.,  
RA Glaser P., Frangoul L., Buchrieser C., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurge O.,  
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
RA Gantier L.-M., Kaerst U., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,  
RT "Comparative genomics of Listeria species."  
RL Science 294:849-852(2001).  
DR EMBL: AL59163; CAC95241.1;  
DR Listlist: LIND0008;  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC: 2.  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 504 AA; 56766 MW; FA019D077F9056EC CRC64;  
SQ

Query Match 11.7%; Score 318; DB 16; Length 504;  
Best Local Similarity 23.6%; Pred. No. 1e-15;  
Matches 123; Conservative 76; Mismatches 213; Indels 110; Gaps 17;

QY 4 NPKTQAMPESETISLMTKRSLSLCLLSCSSWLPLEERTESRHNTSKPVRILNLIQ 63  
DB 82 NPATKRF--STAQVMEKALINAHAIPNNTNEKLPRLSKRI--AHLSIEPIK-GNKIE 136  
QY 64 IRRHTHTGSLDIYLLNDPHEAFARAALIESAHSLDQYIYWRNDISGRLLFNLYLA 123  
DB 137 IL---TNG-----EETFPVLLDLAKRAENHIHQYIYIKTDAISTEIRILVEK 182  
QY 124 AERGVAVLLDDNNTRGDILLALDSHPNIEVLFNPFVLRKWRALGYLDFPRLNRR 183  
DB 183 AKSGVEVRMPFGGLSSKSLKAFAPLKEAGVSIHAFD-IASPIV-----RTANLR 234  
QY 184 MHNKSFADNRTATLIGRNIGDEY-----FKVGEDTVFADDLIATGSVGEVSHDF-- 235  
DB 235 NHRKIVIDQIGTGGNLIGEYRSNPPDFRWRDT-----HMKITGAVIELOESFLN 289  
QY 236 DRYAASHSHANATRIIRSGNIGKGLQALGYNDTSRHALRYRETVESQPLYOKIQTGRI 295  
DB 290 DWIYMENAGAADQFISEAGLQOYFSPVDVGDMA-----QVIYGGPY 332  
QY 296 DMQSVQTRLISDDPAKGLDRDRKRPPIAGRLQDALKEPKSVYLVSPFVPTKSGTDALA 355  
DB 333 D-----KERWVRDSMLDLIDSAESVWIVSPFVPEALAVIR 371  
QY 356 KLVGDIDVTVLNLSQATVAHVHSGYKYPKPLKAGIKLYEOPNNAVATKDKGLT 415  
DB 372 RVAMSGVAVRVIIIG--KGDRGISFHGSNAVYKTMIEAGAKMYATADDSFV----- 420

QY 416 GSSVTSLHAKTFYDGRKIFIGSFNLDPRSARLNTMGVIESPKIAOMERTLATDSPE 475  
DB 421 -----HAKAMLVDTGTRAIAGTANFDVRSFRLNHELVFLUDE--SEAMHLKDKDFKD 471

QY 476 YAVRYTLDRHNLQWHDPAITKTYNPEAKIMKRIAAKIIS 517  
DB 472 F-----EDSRLL-----FTMKDMENKPLLTRIKREVSLSLS 501

RESULT 14  
Q8YAV5 PRELIMINARY; PRT; 504 AA.  
AC 08YAV5;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical protein lmo0008.  
GN LMO0008.  
OS Listeria monocytogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
OC Listeriaceae; Listeria.  
OX NCBI\_TaxID=1639;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-EGD-E / SEROVAR 1/24;  
RX MEDLINE=21537279; PubMed=11679669;  
RA Baguero F., Frangoul L., Buchrieser C., Rusniok C., Amend A.,  
RA Glaser P., Frangoul L., Buchrieser C., Brandt P., Chakraborty T.,  
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deloux P.,  
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurge O.,  
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,  
RA Gantier L.-M., Kaerst U., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
RA Jones L.-M., Kaerst U., Krefit J., Kuhn M., Kunst F., Kurapkat G.,  
RA Madueno E., Maitournan A., Mata Vicente J., Ng E., Nedjari H.,  
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.,  
RT "Comparative genomics of Listeria species."  
RL Science 294:849-852(2001).  
DR EMBL: AL591973; CAC98223.1;  
DR Listlist: LMO00008;  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDC: 2.  
DR SMART: SM00155; PLDC: 2.  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 504 AA; 56708 MW; B9BA4030CE06444D CRC64;  
SQ

Query Match 11.5%; Score 311; DB 16; Length 504;  
Best Local Similarity 23.6%; Pred. No. 3.4e-15;  
Matches 123; Conservative 72; Mismatches 217; Indels 110; Gaps 17;

QY 4 NPKTQAMPESETISLMTKRSLSLCLLSCSSWLPLEERTESRHNTSKPVRILNLIQ 63  
DB 82 NPATKRF--STAQVMEKALINAHAIPNNTNEKLPRLSKRI--AHLSIEPIK-GNKIE 136  
QY 64 IRRHTHTGSLDIYLLNDPHEAFARAALIESAHSLDQYIYWRNDISGRLLFNLYLA 123  
DB 137 IL---TNG-----EETFPVLLDLAKRAENHIHQYIYIKTDAISTEIRILVEK 182  
QY 124 AERGVAVLLDDNNTRGDILLALDSHPNIEVLFNPFVLRKWRALGYLDFPRLNRR 183  
DB 183 AKSGVEVRMPFGGLSSKSLKAFAPLKEAGVSIHAFD-IASPIV-----RTANLR 234  
QY 184 MHNKSFADNRTATLIGRNIGDEY-----FKVGEDTVFADDLIATGSVGEVSHDF-- 237  
DB 235 NHRKIVIDQIGTGGNLIGEYRSNPPDFRWRDT-----HMKITGAVIELOESFLN 289  
QY 238 YVA--SHSAHANATRIIRSGNIGKGLQALGYNDTSRHALRYRETVESQPLYOKIQTGRI 295  
DB 290 DWIYMENAGAADQFISEAGLQOYFSPVDVGDMA-----QVIYGGPY 332  
QY 296 DMQSVQTRLISDDPAKGLDRDRKRPPIAGRLQDALKEPKSVYLVSPFVPTKSGTDALA 355

```

Db 333 D-----KMKVWDSDMLDIDSAKESWIYSPYFPDESLAVIR 371
Oy 356 KLVQDGIQVTLNLSQATDVAHVSGYKVRKPKLLKAGIKLYELOPNHNAVPTKRGKT 415
Db 372 RAMSGVDVRAVITPG--KQDRGISFPGSNVAYKTMTEAGAKMVAADDFV-----420
Oy 416 GSSVTSLSAKFTYVDGKRIFGSPNLDPRSARLNTMGVIESPKIAEQMERTLADTSP 475
Db 421 -----HAKMALVDCTRAIGTANFDVSPRLNHELMFVLDE--SEAMHLLKRPDKD 471
Oy 476 VAYRVTLDRHNRLOHMDPATRKTYNEPEAKLMKRIATILS 517
Db 472 F-----EDSRU-----FTMKDMENKPLLTRKEVLSILS 501

```

## RESULT 15

```

O99SG9 PRELIMINARY: PRT: 494 AA.
ID O99SG9
AC O99SG9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein SAV2088.
GN SAV2088 OR SA1891.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iano J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Murakami H., Hoshiyama A.,
RA Mizutani-U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuuchi J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus."
RT Lancet 357:1225-1240(2001).
DR EMBL: AP003364; BAB58250.1;
DR EMBL: AP003136; BAB43175.1;
DR InterPro: IPR000209; Peptidase_S8.
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDc; 2.
DR SMART: SM00155; PLDc; 2.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
KM Hypothetical protein; Complete proteome.
SO SEQUENCE 494 AA; 56518 MW; 81A5D9B7F098990B CRC64;

```

```

Query Match 11.4%; Score 308.5; DB 16; Length 494;
Best Local Similarity 24.5%; Pred. No. 5, 2e-15;
Matches 107; Conservative 65; Mismatches 143; Indels 122; Gaps 14;

```

```

Oy 74 SDIYLLNDPHEFAARAALIESHSLDQYIYWRNDISGRLLFNLYLAEGVAVRLL 133
Db 131 NDKRITTDQEKFDLIDIRNATDIHQYIYIODELGRITLNLGKAEQGVKTL 190
Oy 134 LDDNNTRGDLDDLALDSHPNIEVRLFPFLKMKWALGYLTDFP-----RLNRNH 185
Db 191 YDDMGSRGL-----RKKGRLRPFRRKNGHAEP--FPSKPLPILNRNRRNH 224
Oy 186 NKSFTADNRATILGRNIGDEYFKVEDTVFA---DLDIATGVSVEGSHDFDRYWSH 242
Db 235 RKLIVIDIGQIGYGVNGVDEY--LGSKSKFGYWRDTHLRIVGDVAVNALQRLTDMNS- 291
Oy 243 SAHNAIR-----IIRSGNIGKGLQALGYNDESRALLKXREIVEOSPXYOKI 290
Db 292 ---QATRDHISYDRYFPDVSNGTIGVOIASSGPDE-----WEOL 330

```

```

Oy 291 QTRIDMQSVQTRRLISDDBAKGLDRDRKRPPIAGRLQDALKQPEKSVLYVSPYFTKSG 350
Db 331 KYGYL-----KMIS-----SAKSITIQSPYFPDDQAF 358
Oy 351 TDALAKLVODGIDVYTLNLSQATDVAHVSGYKVRKPKLLKAGIKLYELOPNHNAVPTK 410
Db 359 LBSIKIALGALGVDMINEN--KPDHPFVFWATLKNAAASLLDAGYKVFHY-----406
Oy 411 DKGLGSSVTSLSAKFTYVDGKRIFGSPNLDPRSARLNTMGVIESPKIAEQMERTLA 470
Db 407 DNGF-----LHSLKTVLIDDEIASVGTANMDHRSFTLNEVNAFYDQOIAKKLQAFI 459
Oy 471 D-----TSPEYAVR 479
Db 460 DLLAVSSELTAKARYAKR 476

```

## RESULT 16

```

O9CGY2 PRELIMINARY: PRT: 481 AA.
ID O9CGY2
AC O9CGY2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Cardiolipin synthase.
GN CLSA OR LD0960.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11403;
RX MEDLINE=21233186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jallion O., Malarme K.,
RA Welschenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis 11403."
RT Genome Res. 11:731-753(2001).
DR EMBL: AE006330; AAK05058.1;
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDc; 2.
DR SMART: SM00155; PLDc; 2.
KM Complete proteome.
SO SEQUENCE 481 AA; 55825 MW; 80B3621FC6343C0 CRC64;

```

```

Query Match 10.8%; Score 293.5; DB 16; Length 481;
Best Local Similarity 22.6%; Pred. No. 6, 9e-14;
Matches 110; Conservative 83; Mismatches 171; Indels 123; Gaps 17;

```

```

Oy 54 KPVRLDNL-QIRHPTH-----TNGLSDIYLLNDPHEFAARAALIESHSLDQY 104
Db 92 KKRQNHVIGQLHMLYVEEKSTISTN--TGKIFNDGKRFALIDDINKAKHHVHMEY 149
Oy 105 YWRNDISGRLLFNLYLAERGVAVRLLD--DNNTGDLDDLALDSHPNIEVRLFN 161
Db 150 YIFRDMRGMHEIYDALLAAKQGVYKLLIDMGSKTK-----MNFKELO-----196
Oy 162 PVLRRKRALGYLTDF-----PRLNRMHNSFTADNRATILGRNIGDEYFKVEE 212
Db 197 -----EAGHVAQFEPPLIMPLVNRPTVRLRKRIIVIDIGLIGYGVNGVDEYASITK 249
Oy 213 D-TVEADLDILATGVSVEGSHDFDYKASHSHNNATRIIRSGNIGKGLQALGYNDETSR 271
Db 250 KEGYWRDNLRLTGLDIVYSLQHRFTLDW-----NSQ 280
Oy 272 HALLRRETVESPLY-QKIQTRIDMQSVQTRRLISDDBAKGLDRDRKRPPIAGRLQDAL 330
Db 281 H-----HFVYEGEGRFPDSIVESH-----VATQLVISGP-----DEDEQIKLYTMKMI 325
Oy 331 KQPEKSVLYVSPYFTKSGTDALAKLVODGIDVYTLNLSQATDVAHVSGYKVRKPKL 390

```

```

Db 326 SGAEEIIITQPIYPYIPSPALHESIKLALISGVQVKKLIPN--KPDHPLYWATYTHAADL 383
QY 391 LKAGIKLYELOPNHNAVPAKDKGLTSSVSTSLHAKFTIVDGKRIFFSGNLDPRSARLNT 450
Db 384 VKYGAKVY-----TYENGK-----VHSKTLIIDGYSVGSANLIDRSLQLOCF 426
QY 451 EMGVYIESPKIAEOMERLADTSPRYATRVTLDRHNRLOMHPDPAIRKTYPNPEPKLKKR 510
Db 427 EANNVYIYDYSQKLRN-----DFMKDKLSRPLRLRYERSKSLVRPE 471
QY 511 IAAKILS 517
Db 472 GLARLIA 478

RESULT 17
Q92720 PRELIMINARY: PRT: 482 AA.
ID 092720
AC 092720:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein lin2646.
GN LIN2646.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Chabalt A., Cheounani F., Couve E., de Daruvar A., Deloux P.,
RA Domant E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fathi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kunz M., Kunst F., Kutuphat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstjek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.".
RL Science 294:849-852(2001).
DR EMBL: AL596173; CAC97873.1;
DR Listlist; LIN02646;
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDc: 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 482 AA; 55462 MW; AOA3F18B99922EC7 CRC64;

Query Match 10.8%; Score 293.5; DB 16; Length 482;
Best Local Similarity 24.6%; Pred. No. 6.9e-14;
Matches 113; Conservative 60; Mismatches 172; Indels 115; Gaps 16;

QY 74 SDIYLLNDPHEAPARAALISAHSIDLQYIMRNDISGRLENLVYLAERGVRL 133
Db 119 NEVELFIDGHEKFDALIDIEKAKDHILYIFHSDELGRIMLRVBRKAEGCLNVKII 178
QY 134 LDDNNTRGDLDDLALDSDHPNIEVRLNPFVLRKRAALGYLDPRLRRHNSFTADN 193
Db 179 YDAMGSRRTTKSFRTTEKNGCLVAPFPSPKLP-----LINF-RLNVRNRKLAITIDG 230
QY 194 RATLLGGRNIDEGYFKVEDVAFDLIDLATGVSVEVSHDFRWASHASHNATRIIRS 253
Db 231 DVGVIIGFNIDEX--LGRSKKFG-----YWRD-----THLAVH 262
QY 254 GNIGKGLQA---LGYNETSRHAL---LRYRETV--EOSPLYOKIQTG-RIDWOSVQTRL 304
Db 263 GKAVAMQTRFRTIMDNMSASTHKKIDYKARYPPTFGKCHTSMQIVSSGPDSEMOQIKNGY 322
QY 305 ISDDPAKDLDDRRKRPPIAGRLQDALAKQPEKSVYLVSPYFPTKSGTALAKLVQDGDIV 364

```

```

Db 323 IK-----MINAKKTIYLOSYPFIPDASLEAIRIAALSGVDV 360
QY 365 TVLTNSLOATDVAVAHSGVYKRPKLKAGIKLYELOPNHNAVPAKDKGLTSSVSTSLHA 424
Db 361 RVMIPN--KPDHAFYRRTTYNAGELMETGAKIF-----IYDNGF-----IHA 401
QY 425 KTFIVDGKRIFFSGNLDPRSARLNTENGVIIESPKIAEOMERLADTSPRYATRVTLDR 484
Db 402 KTLVYDGEIASVGTANMDFSRFLNFEVNAFYERKKMVKQLEDAFLD----- 449
QY 485 HNRLOMHPDPAIRKTYPNPE---AKLM---KRIAKILS 517
Db 450 -----ILKSYQLPELYAKRSIMWIKERKAVSRLLS 479

RESULT 18
Q8XP94 PRELIMINARY: PRT: 480 AA.
ID Q8XP94
AC Q8XP94:
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable cardiolipin synthase.
GN CLSA OR CPE0071.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / TYPE A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flash-eater.".
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL: AP003185; BAB79777.1;
DR InterPro: IPR001736; PLD.
DR Pfam: PF00614; PLDc: 2.
DR SMART: SM00155; PLDc: 2.
KW Complete proteome.
SQ SEQUENCE 480 AA; 55597 MW; 30BA57D10662D792 CRC64;

Query Match 10.8%; Score 292.5; DB 16; Length 480;
Best Local Similarity 22.3%; Pred. No. 8.2e-14;
Matches 105; Conservative 85; Mismatches 185; Indels 95; Gaps 13;

QY 39 PLEERTSRH-----FNTSKFVRLDNLQIRHPTHTGLSDIYLLNDPHEAPARA 90
Db 78 PILDEYPERKQYVPLMELLAFNSDSPIFVNN-----DIKIKGVKGFYDK 123
QY 91 ALISAHSIDLQYIMRNDISGRLENLVYLAERGVRLDLDNNTRGDLDDLALD 150
Db 124 EELKAKAHNHLEVIYNSDEIGNEIKDILKKALEGVKVFITDKYSSSLKASYKDL 183
QY 151 SHPNIEVRLNPFVLRKRAALGYLDPRLRRHNSFTADNRTATILGGRNIDEGYFKV 210
Db 184 KKAGISVIMSYFLAPLKLVIN-----TQIVNRNRHRLIVIDGQVGLGGINIDEX--L 236
QY 211 GEDVYFADLIDLATGVSVEVSHDFRWASHASHNATRIIRSNGIKGLQALQYND--- 267
Db 237 GRNKKFG-----YWRD-----THIMIKDFFVLAQAVFLDFFIT 270
QY 268 -ETSRHALRYRETVESPLQOKIQTGRIDWOSVQTRLISDDPAKGLDROKRPPIAGRL 326
Db 271 IERANNSTFYDKEDP-----KYRPENIVAKERYLMQLVKSGP-----DSTFPAIMQSV 319
QY 327 QDALQPEKSVYLVSPYFPTKSGTALAKLVQDGDIVTVLTNSLOATDVAVAHSGVYK 386
Db 320 LKMITMARENITITTPYVPPSSITIEALRIASLSGVQDKIIFP--EKSDHFMVAKASKSY 377

```

QY 387 RRPLKAGIKIYELQHNHAPATKDKGLTGSVTSIAKTFIVDGKRIFFSGFNDPRS 446  
 DB 378 LALMOCGIEV-----FYDKS-----SFTHSKTMTIDGICILGTANMIRSF 421  
 QY 447 RLNTGCVVIESPKIAEQMERIADTSPEYAVRYTLDRHNLQWMDPATR 496  
 DB 422 ELNVEINFVIYDKETCKLDALFEED-----LTKSRHFKIEYEKSTK 464

## RESULT 19

Q8XMD5 PRELIMINARY; PRT; 485 AA.  
 AC Q8XMD5;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN CLSC OR CPE0754.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostridiales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxId=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-13 / TYPE A;  
 RX PubMed-11792842;  
 RA Shultz T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic  
 RT flesh-eater."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL; AP003188; BAB80460.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART; SM00155; PLDC; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 485 AA; 56193 MW; B988569756D97CD6 CRC64;

Query Match 10.7%; Score 289.5; DB 16; Length 485;  
 Best Local Similarity 23.4%; Pred. No. 14e-13;  
 Matches 115; Conservative 90; Mismatches 176; Indels 109; Gaps 20;

QY 42 EERTSRHNTSKPVRLLDILQIRHPHTNGLSIDYLLNDPHEAFARALIESAHSID 101  
 DB 88 DDLKRAKSHDMKIAIE-----SMEYSPIKDY-ETKVFPGKELFDLIESLKAKQSIN 141  
 QY 102 LQYITRNDISGRLEFNLYLAERGVRYRLDLD--NNTRGIDLLLDLSDHPNIEVRL 159  
 DB 142 IEFYIKNDIDIGSKVLDILKEKAKSGEVRLLYDSVGSRTNKKLQSAIDA--SVKGE 199  
 QY 160 FNFVLRKRRALGYLDFPRLNRHMKSFPTADNRATILGGRNIGDEYFVGSDTYEA-- 217  
 DB 200 FFFSLR-----LINT-NMNRNRKIIIVIDNKKGYGVGFVGDY--LGADPKFGY 249  
 QY 218 -DDILATGSGVEVSHDFD-RYMAASHANATRIIRSGNIGKQALGYNDSTRHALL 275  
 DB 250 RQTHI-----KMGDSVRDLDLRFEMADM--RYATK-----EDIDISHLI 286  
 QY 276 RYRETEVQSPLYOKIOTGRIDMOSVOTRLISDPAKGLDRDRKP---PIARLDALK 331  
 DB 287 DYKN--EQAP-----STMANYSKCGMOIISGP-----NPNFEYIKLSTYEMIO 329  
 QY 332 QPEKSVYLVSPYFVPTKSGDALAKLVODGIDVTLTNSLOATDVAVHSGYKRPPL 391  
 DB 330 KAKKIYIIGSPYLLDINSISDLSKLASISGVDRIMPG--KQDHPFYVANIKNYAGDLL 387  
 QY 392 KAKIKYLELOHNHAPATKDKGLTGSVTSIAKTFIVDGKRIFFSGFNDPRSARLNT 451  
 DB 388 NFGVKIYHDKN-----AFHSGKTLVIDEVCISIGANMDTSHFELNFE 431  
 QY 452 MGAVIESPKIA-----EQMERTIADTSPEYAVRYTLDRHNLQWMDPATRKTYYNEPAKL 507  
 DB 432 VIAMLYSDEIAKIQROFEDMLISK-----ELTRDYLAKGTSTVK 472

QY 508 WKRIAKILSL 519  
 DB 473 IKESFKLSAL 484

## RESULT 20

Q8YAE3 PRELIMINARY; PRT; 482 AA.  
 AC Q8YAE3;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical protein lmo2503.  
 GN LMO2503.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Listeriaceae; Listeria.  
 OX NCBI\_TaxId=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EGD-E / SEROVAR 1/2A;  
 RX MEDLINE-21537279; PubMed-11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitounam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordiel G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
 RT "Comparative genomics of Listeria species."  
 RL Science 294:849-852(2001).  
 DR EMBL; AL591983; CAD00581.1; -  
 DR ListList; LMO02503; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART; SM00155; PLDC; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 482 AA; 55362 MW; D47B2CDBE14651A8 CRC64;

Query Match 10.6%; Score 288.5; DB 16; Length 482;  
 Best Local Similarity 24.3%; Pred. No. 1.7e-13;  
 Matches 112; Conservative 61; Mismatches 172; Indels 115; Gaps 16;

QY 74 SDIYLLNDPHEAFARALIESAHSIDLQYITRNDISGRLEFNLYLAERGVRYRL 133  
 DB 119 NEVELFVDGHEKEDALADIIEKAKDHILYIFHSDLEGNRLMRYLERKAAGLVKII 178  
 QY 134 LDDNNTRGIDLLLDLSDHPNIEVRLFNPNVLRKRRALGYLTPPRLNRHMKSFADN 193  
 DB 179 YDMGSGTTRKSPFRFGKNGLVRFEPFSKL-----LINF-RLVNRRHKKLIIDG 230  
 QY 194 RATILGGRNIGDEYFVGEDTFPADDILATGSGVGEVSHDFDPRYMAASHANATRIIRS 253  
 DB 231 DVGITGSGFNIGDEY--LGASKKFG-----YMRD-----THLRVH 262  
 QY 254 GNIGKGLQA--LGYNDSTRHALL--LRYRETV--EQSPLYOKIOTG-RIDMOSVOTRL 304  
 DB 263 GRAVYVMQTRFIMDMNSASTHKIDKARVFPFHGKSHSMQIVSSGDPSEMOQIKNGY 322  
 QY 305 ISDDPAKGLDRDRKRPPIAGRLDALKQPKSVYLVSPYFVPTKSGDALAKLVODGIDV 364  
 DB 323 IK-----MINAAKKTIVLOSPEYIPDASLEAKIAKIALSGVDV 360  
 QY 365 TVLTNSLOATDVAVHSGYKRRPKLAKIKIYELQHNHAPATKDKGLTGSVTSIAH 424  
 DB 361 RVMPIN--KPDHAFVYRATINIVAGELMETGAKIF-----YDNGF-----IHA 401  
 QY 425 KTFIVDGKRIFFSGFNDPRSARLNTMGVYIESPKIAEQMERIADTSPEYAVRYTLDR 484

```

Db 402 KTLVVDGELASVGTAMMDRFRSLNEFVNAFLYKQWVKLDALEDD----- 449
Qy 485 HNRLOMHPATKRTYPNPEPE-----AKLW---KRIAAKILS 517
Db 450 -----ILKSYQLPEPELAKRSIMWIKFEAVSRLLS 479

RESULT 21
Q8XQ0L1 PRELIMINARY; PRT; 489 AA.
ID Q8XQ0L1
AC Q8XQ0L1
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Probable cardiolipin synthetase transmembrane protein
DE (EC 2.7.8.-).
GN CLS OR RSP1211 OR RS03155.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
RA Sliuier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646083; CAD18362.1;
DR InterPro; IPR001736; PLD.
DR Pfam; PF00614; PLDc. 2.
DR SMART; SM00155; PLDc. 2.
KW Transferase; Plasmid; Complete proteome.
SQ SEQUENCE 489 AA; 54855 MW; 801F827E9D40C573 CRC64;

Query Match 10.6%; Score 286.5; DB 16; Length 489;
Best Local Similarity 24.6%; Pred. No. 2.4e-13;
Matches 108; Conservative 59; Mismatches 181; Indels 91; Gaps 12;

Qy 68 PHTNGL-----SDIYLLNDPHEAFARAALIESAHSLSLOYIWRNDISGRLLFNLYL 122
Db 119 PRLTGMPLARRRRLVLYNGAETFAIKARAOVLFEQFIVHDALGRRLALELLE 178
Qy 123 AAERGVARRLLDDNNRTGGLDLDLALDLSHPNIEVRLFNPFVLRKWRALGYLTDFPRLNR 182
Db 179 RARAAVRITYLEFDRIQSALPGHYVROLGAGVWVRPESTHA-----GFVNR-QLNF 230
Qy 183 RMHNSFTADRNRATILGGRNIGDEYKVEDYVFA---DLIOLATGSVYGEVSHDF--DR 237
Db 231 RNRKLVVVDGERAVGGHNVTET--LGEHPPLAPWRDTHIEIYGAAVMDLQTLFAEDW 288
Qy 238 YWASHSAHNAATRIIRSGNIGKGLQALGYNDETSRHALLRYRETYEVSPLYOKIQTGRIW 297
Db 289 YWAHEVQ-----LIVPRRGPEDMWCQVVASGPRDP 321
Qy 298 QSVQTRLLSDPAKGLDRKKRPIAGRLQALAKOPEKSVYLVSPYFVPTSGDALAKL 357
Db 322 QETCSLF-----FVEAIGARRKRLMTTPYFIPEDEAVFALRLA 360
Qy 358 VQDGDIVVLTNSLOATDVAAVHSGVYKRPFLKAGIKLYELOPNHNAVPAKDKGLTGS 417
Db 361 VLRGVDVILLPA--RPDHRVVFHASTLYAVQAIRAGVAKIRYILGF----- 405
Qy 418 SVTSLHAKTFIVDGKRIFGSFNLDPRSARLNTGEMGVVIESPKIAEQMERTLADTSPEYA 477
Db 406 ----LHQKVVLLDDEAAVAGSANLDNRSFRLNEFLMVTADHRFANDVAQMLEADFAF-A 460

```

```

Qy 478 YRVTLDRHNRLOMHPATR 496
Db 461 TRIGDEYEHAN---PARR 476

RESULT 22
Q8R663 PRELIMINARY; PRT; 479 AA.
ID Q8R663
AC Q8R663
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cardiolipin synthetase (EC 2.7.8.-).
GN FN0923.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=21886394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Forstein M., Kypides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586.";
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL; AE010600; AAL95119.1;
KW Transferase; Complete proteome.
SQ SEQUENCE 479 AA; 56509 MW; 5C4CB432BF7CEFF70 CRC64;

Query Match 10.4%; Score 283; DB 16; Length 479;
Best Local Similarity 22.5%; Pred. No. 4.3e-13;
Matches 105; Conservative 83; Mismatches 180; Indels 98; Gaps 14;

Qy 62 LQIRTPHTNGLSDIYLLNDPHEAFARAALIESAHSLSLOYIWRNDISGRLLFNLY 121
Db 105 LEMSTDNISSNNDIQVFTGEDPEPELKEIANAKKRFNMEYFQDGIQKEIADLLI 164
Qy 122 LAERGVARRLLDDNNRTGGLDLDLALDLSHPNIEVRLFNPFVLRKWRALGYLTDFP--- 178
Db 165 EKAKGEVENLILDGVLNANF--RLKQYFKNGVNLHLF-----FRYIPIFN 210
Qy 179 -RLNRMHNSFTADRNRATILGGRNIGDEYKVEDYFADLILATGSVYGEVSHDFR 237
Db 211 IRLNRNRNRKTYIIDNRILAFVGMNIGDEYLGKGIYWRDTSVKIYDIYSFPEKEF-- 268
Qy 238 YWASHSAHNAATRIIRSGNIGKGLQALGYNDETSRHALLRYRETYEVSPLYOKIQTGRIW 297
Db 269 YFSLISIVN-----EFLKDEKFSNLS--LKYEED----- 296
Qy 298 QSVQTRLLSDPAKGLDRKKRPIAGRLQALAKOPEKSVYLVSPYFVPTSGDALAKL 357
Db 297 EGIYQMLISSG-----NYEPALRDVYIKLIQEARSVFIQPYFEPDOLLDTLTKSA 350
Qy 358 VQDGDIVVLTNSLOATDVAAVHSGVYKRPFLKAGIKLYELOPNHNAVPAKDKGLTGS 416
Db 351 VLSGIDVAKIMIPNKRADHPFIWVNOYV---WELLRLCANIYRYNGF----- 395
Qy 417 SSVTSLHAKTFIVDGKRIFGSFNLDPRSARLNTGEMGVVIESPKIAEQMERTLADTSPEY 476
Db 396 ----IHSKTLVDEEVVSGVGNDFYSPFLNFEINININKEVANSFK-----AQY 443
Qy 477 AYRVTLDRHNRLOMHPATRKYTPNEPEAKLMKRIAAIISLL-PI 521
Db 444 YKDITISK-----KLTFADFKKRSIFTYKESVFRLLSPI 478

RESULT 23
Q8RK18

```



ID O8RK18 PRELIMINARY: PRT; 481 AA.  
 AC O8RK18;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN CLS.  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas;  
 OX NCBI\_TaxID=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=P8;  
 RA von Wallbrunn A., Heipleper H.J., Meinhardt F.;  
 RT "cis/trans isomerisation of unsaturated fatty acids is not affected in  
 a cardiolipin synthase knock-out mutant of P. putida P8."  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBD databases.  
 DR EMBL; AJ441309; CAD2690.1;  
 SQ SEQUENCE 481 AA; 54396 MW; 627AD6D566833FC6 CRC64;

Query Match 10.4%; Score 281.5; DB 2; Length 481;  
 Best Local Similarity 24.4%; Pred. No. 5,6e-13;  
 Matches 122; Conservative 68; Mismatches 190; Indels 119; Gaps 16;

OY 36 SMLPPEERTESRHEFTSKPRVLDNIQIRHTPTNGL-----SDIYLLNDPHEAFARA 90  
 DB 87 NMRPWEELTAR-----ESESVALRAMKLGMRCLANNQVKLLIDGKATPDALF 138  
 OY 91 ALIESAHSLDQYYIWRNDISGRLLFNLYLAERGVRRLDDNNTRGLDILLALD 150  
 DB 139 AADIKARADVLYOFFIIHDDTLGKALQQLLRKAAGVOVFVLYDRIGSHALPARYSOOL 198  
 OY 151 SHPNIEVRLEFNPVLAKKMAGLGLDPRLL--NRMHNSFTADNKAITIGSGNIDDEF 208  
 DB 199 RGGVOIHAF--ATRRGW-----FNRLQVLRNHRKIIVVDGLGLGIGHNVDGY- 247  
 OY 209 KYGED--TFADLDILATGSVYGEVSHDF--DRYWASHAHNATRIIRSGNIGKLOAL 263  
 DB 248 -LGEHKKLPWRNTHQISGPVLACQESFAEDWYAT-----RQLPRL 290  
 OY 264 GINDETSRHALLRYRETVESQSLYOKIOTGRIDWOSVOTRLISDDPAKGLDRDRKPPA 333  
 DB 291 ILPDYPR-----ONGVLCQALASGRADPOETCSLF----- 320  
 OY 324 GRLODLAKQPEKSVLYSPYFPTKSGDALAKLYODGIDVYLTNSLOATDVAHVSGY 383  
 DB 321 --PLEAHSATRKWITSPFIDPAVFALRLAVLRGVDRVLPS--RPDRVYVAAS 376  
 OY 384 VKYRKPLKAGIKLYELOPNHNAVPAKDKGLTSSVLSHAKTFIVDGRIEIGSFNLDP 443  
 DB 377 SLFAFAVAGAGIMFRYQGF-----LHQKVVLDVDDDSAGISPRLDN 419  
 OY 444 RSRRLTEMGVLTSPKIAEQMERTLA--DISPEYAVRTLDNRNLRQMDPATRKYTP 500  
 DB 420 RSRIRLETLLVDRRPAFOYEKMLINDEQSRITAEDESODTH--RLQ----- 467  
 OY 501 NEPEAKMKRIAAKILSL 519  
 DB 468 -----QLGMRIARLISPIL 481

RESULT 24  
 O99UF8 PRELIMINARY: PRT; 505 AA.  
 ID O99UF8;  
 AC O99UF8;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE SAIL155 protein (Cardiolipin synthetase homolog).  
 GN SAV1317 OR SAIL155.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and  
 OS Staphylococcus aureus (strain N315).

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OC Staphylococcus;  
 OX NCBI\_TaxID=156878, 156879;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-T., Nagai Y., Itoh T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 aureus."  
 RL Lancet 357:1225-1240(2001).  
 DR EMBL; AP003362; BAB57479.1;  
 DR EMBL; AP003133; BAB42411.1;  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; PLDC; 2.  
 DR SMART; SM00155; PLDC; 2.  
 KW Complete Proteome.  
 SQ SEQUENCE 505 AA; 57757 MW; 32B4DACA73EE01E0 CRC64;

Query Match 10.1%; Score 275; DB 16; Length 505;  
 Best Local Similarity 21.5%; Pred. No. 1.9e-12;  
 Matches 100; Conservative 81; Mismatches 149; Indels 136; Gaps 16;

OY 74 SDIY--LNDPHEAFARAALIESAHSLDQYYIWRNDISGRLLFNLYLAERGVRN 131  
 DB 151 NDLYDGLKD-----IKNAKEYIHLEYTFALDGLGRILHALEEKKGLEEV 199  
 OY 132 LLLDDNNTRGLDILLALDSHPNIEVRLEFNPVLAKKMAGL-----YLDP--RLN 181  
 DB 200 ILYDVGSK-----NVKMANPDHF-----KSLGGEVEAFASKPLRLNFRM 241  
 OY 182 RRMHNSFTADNKAITIGSGNIDDEFKYGEIDVFADLDILATGSVYGEVSHDFRYMAS 241  
 DB 242 NRRHKKIIVIDGOLGVYGGFNIDGYLGCKLQKWMDBTHIRIGDANVALQALFIIDWNS 301  
 OY 242 HSNHNRTRII-----RSGNICKGLQALGYNDETSRHALLRYRETVESQSLYOKIOTGR 294  
 DB 302 -QAHRPQFEYDVKYFFPKKNPRLG-----NSPIQIAASGPA 335  
 OY 295 IDMGVQ--TRLSDDPAKGLDRDRKRPPIAGRLQDALQPEKSVLYSPYFPTKSGT 351  
 DB 336 SDNHQIEGYTKMI-----MSAKSVYLOSYPYFIPDYSYI 370  
 OY 352 DALAKLVODIDVYLTNSLOATDVAHVSGYVYKRPPLKAGIKLYELOPNHNAVPAKTD 411  
 DB 371 NAIKIAKSGVDNHLMPG--KPDHPLVYMATPSNMSDLSGVKII-----TYE 418  
 OY 412 KGLTGSVVSLSHAKTFIVDGRIEIGSFNLDPRSARLNTMGVLTSPKIAEQMERTLAD 471  
 DB 419 NGF-----IHSKMCILIDEIVSGFANMDFSFELNFEVNFVYDENLAKDL----- 465  
 OY 472 TSPDYAVRTLDNRNLRQMDPATRKYTPNEPEAKMKRIAAKILS 517  
 DB 466 ---RVAEHDITKSKOL-----TKESYANRPVSVAFKESLAKLVS 502

RESULT 25  
 O9HTHO PRELIMINARY: PRT; 490 AA.  
 ID O9HTHO;  
 AC O9HTHO;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN CLS OR PA5394.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;





GN PA4339.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.B., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
Hickey M.J., Brickman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
Gardner R.L., Goltz L., Tolentino E., Westbrook-Wedman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Ladbeg K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL: AE004850; AAC07727.1;  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDc: 2.  
DR SMART: SM00155; PLDc: 2.  
KW Complete proteome.  
SQ SEQUENCE 359 AA; 41408 MW; E53EE2DA7B1D8426 CRC64;  
Query Match 9.6%; Score 261; DB 16; Length 359;  
Best Local Similarity 24.1%; Pred. No. 1.3e-11;  
Matches 102; Conservative 58; Mismatches 161; Indels 102; Gaps 14;  
QY 93 IESAEHSIDLOYIWRNDISGRLEFNLVYLAERGVRRLDDNNRTGIDLLALDSH 152  
DB 5 IEAORSIELELYVEDGHCAELFLVALLDARRRGVAVRCLEDFGCGLSAWIQRLRE 64  
QY 153 PNIERLNPFVLKRWALG---YITDEPRLNRHMKSFADNRATILGCRNIGDEYFK 209  
DB 65 AGGELRLNPL--RWKLTGMLY-----RDHRLLLVDGRLGYVGAGITDEFWE 112  
QY 210 -VGEITYVADLILATGSGVGEVSHDFRYVASHSAHNA-----TRIRSGNIGKG 259  
DB 113 PVSIDSANREVMEDGVVADWALFEROMIACLEEKAMKPREMTILTRLPQGAAR 172  
QY 260 LQALGVNDSTSHALLRREYVESQSPLYQIQTGRIDWQSVOTRLISDDPAKGLDRRRK 319  
DB 173 LGRVAYADARQHRDL-----QSL-----VALNGSRRR 201  
QY 320 PPIARLQDALKEPKSVYVSPYVPTKSGDALAKLVQDGDIVYVLTNSIQATIDVAAY 379  
DB 202 -----TWLATPYFLPTWKVRRALRKAARGVEVRLLAG--RLTDHAPV 243  
QY 380 HSGYKVKRPLLKAGIKIKEYLQPNHNAVPAATKDKGLTGSSVTSLHAKFTLVDSKRIFFISG 439  
DB 244 RYAGGRYTPRLLRAGVRRIHETQPR-----LHLKMWVD--DWVSVGSC 285  
Y 440 NLDPASRLNTEMGVVISPKIA--EQMERTLADTSPEYAVRYVTLDR-----HNRL--Q 489  
DB 286 NFDHWNLRFNLDANLEADDPPTNEAASLLADFADSREVTLLAMRARPIAMRHLHQRLMG 345  
QY 490 WHD 492  
DB 346 WLD 348  
RESULT 30  
Q97E04 PRELIMINARY; PRT; 510 AA.  
AC Q97E04;  
DT 01-OCT-2001 (TReMBLrel, 18, Created)  
DT 01-OCT-2001 (TReMBLrel, 18, Last sequence update)  
DT 01-DEC-2001 (TReMBLrel, 19, Last annotation update)  
DE Possible cardiolipin synthase (phospholipase D family).  
GN CAC3316.  
OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;

OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Neelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng O.,  
Gibson R., Lee H.W., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,  
Ratetsky R.L., Sabathe F., Doucette-Stam L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.,  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol 183:4823-4838(2001).  
DR EMBL: AE007829; AAK81248.1;  
DR InterPro: IPR001736; PLD.  
DR Pfam: PF00614; PLDc: 2.  
DR SMART: SM00155; PLDc: 2.  
KW Complete proteome.  
SQ SEQUENCE 510 AA; 59260 MW; 13DAE45CFE268FAE CRC64;  
Query Match 9.5%; Score 258; DB 16; Length 510;  
Best Local Similarity 21.4%; Pred. No. 3.8e-11;  
Matches 104; Conservative 75; Mismatches 187; Indels 120; Gaps 16;  
QY 61 ILQIRHTHTNGLSDIYLLN---DPH-----EAFARAALIESAEHSIDQY 104  
DB 117 INEIRNNNETANOSSYIQNYSLEPPYKYAEYFSTGEFTFNEIOLKRAKOYIFLEY 176  
QY 105 YIWRNDISGRLEFNLVYLAERGVRRLDDNNRTGIDLLALDSHPN----- 154  
DB 177 FLIKGVMMNSYQLRLRQKVGEGVEYVIYDD-----IGCFILPNQYHKELEKIG 227  
QY 155 IEVRLFNPFVLKRWALGILDFPRLNRHMKSFADNRATILGCRNIGDEYF--KGED 213  
DB 228 IKCCVFENPLI-----PMVSF--KFNRRDHKRLAVIDGLVGFQGINLSDIYINKYOKY 278  
QY 214 TVEFADLILATGSGVGEVSHDFRYVASHSAHNAATRIIRSGNIGGQALGVNDSTSHA 273  
DB 279 GYWKDTALKVEKKAAMNLSVWFLSMS-----FLR--GIDEDFTFFKNIEKES 327  
QY 274 LKRYREIVEQSPLYQIQTGRIDWQSVOTRLISDDPAKGLDRRRKPIAGRLQDALKEP 333  
DB 328 REGYVQPEPADSPL-----DGEVGEIIMNLIS-----KA 357  
QY 334 EKSIVLVSPYVPTKSGDALAKLVQDGDIVYVLTNSIQATIDVAAYHSGYKRPPLKA 393  
DB 358 TKYVYITPYLVIGNEVNTALTSAAKGVDAVRIITPHI--PDKRIVHSVTSKYKVLIES 415  
QY 394 GKLYEIQPNHNAVPAATKDKGLTGSSVTSLHAKFTLVDSKRIFFISGPNLDPASRLNTEMG 453  
DB 416 GKLYEYMPGF-----THSKTYVCDNEGVVGSINMDPSRLYLHFECCG 458  
QY 454 VIESPKIAEQMERTLADTSPEYAVRYVTLDRHNRLOMDPATTRKTYVPEPAPAKLMKRIA 513  
DB 459 VWMYKTNIVYDIKDFMOTL--DKSKETILEELINKVKW-----STLM--RYVL 503  
QY 514 KILSL 519  
DB 504 RVFAVL 509  
RESULT 31  
Q9FDK4 PRELIMINARY; PRT; 466 AA.  
AC Q9FDK4;  
DT 01-MAR-2001 (TReMBLrel, 16, Created)  
DT 01-MAR-2001 (TReMBLrel, 16, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel, 21, Last annotation update)  
DE Cardiolipin synthase.  
OS Zymomonas mobilis.  
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;  
OX NCBI\_TaxID=542;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-ZM4;  
 RA Lee H.J., Kang H.S.;  
 RT "Sequence analysis of fosmid clone 42D10 of *Zymomonas mobilis* ZM4.";  
 RL Submitted (Dec-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF212041; AAC02166.1;  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; Pldc; 2.  
 DR SMART; SM00155; Pldc; 2.  
 SQ SEQUENCE 466 AA; 52353 MW; 18FF1P45C4A3BD4 CRC64;

Query Match 9.3%; Score 253; DB 2; Length 466;  
 Best Local Similarity 22.7%; Pred. No. 7.9e-11;  
 Matches 121; Conservative 80; Mismatches 214; Indels 118; Gaps 17;

QY 1 MRANPKTQAMPESETSLMKTRSLISLCLL-----CSCSSWLPPEERTES 47  
 DB 27 MRRTP--QAASSWLLIFQPALGLLIMIGQAFPSWRARFANLKSFLE--EQRKLT 82  
 QY 48 RHEFNKSP-----VRLDNIQIRTPHTNGLSDIYLLNDHEAFARAAALIESAHSLDQ 103  
 DB 83 QSVASLPEALADLADLAEENLGHMPALSGNRMLY--TSDQALFDHLADIDAEQHVHLV 141  
 QY 104 YIWMNDISGRLLFNLVLAERGVRRLDDNNTRGIDDLALDSHPNIEVRLNPF 163  
 DB 142 IYIFSDRSGCOIILAAGAAKRGVNCQILIDALSGRSGKQVIOQLKVLGDARALPF 201  
 QY 164 VLKRRALGYLTFPRLNRHMKSFADNRATILGGRNIGDEYFVGEDTVFADLLIA 223  
 DB 202 RLRRRTNG-----DHRNHRKLFIIIDKIGVAGSONIIRMDWPG--VVOQLTVHV 251  
 QY 224 TGSVGEVHDDRYWASHSAHNATRIISGNIGKGLAGNDESRALLRYPT---- 280  
 DB 252 EEPYFEMETPLAIIW-----YESGALPD-----RMRTPHL 284  
 QY 281 ---VEOSPLYQKIQGRIDMOSVOTRLISDPAKGLDRRRKPIAGRLDALKOPEKSV 337  
 DB 285 VPKESGSHLQLLPSS-----AYRLAGFPTMLVTGIYAAKQOI 323  
 QY 338 YIVSPFYVTKSGTDALALVODGIDVYVINSLOATVAAVHSGVYKRPYLKAGITL 397  
 DB 324 VLVTPPLLPDSLLAKTAAALRGVRIIDIVSKI--ADQILVSMQKSYQELLSAGIH 381  
 QY 398 YELQNHAVPATKDKGLTSSVTSLSHAKTFIVDGKRIFGFSGNLDPRSARLWEMGVVE 457  
 DB 382 HRY-----RDK-----LHAKHISVDNKLAWGSSVNDIRSLNEEVSILTY 424  
 QY 458 SPKIAQEMERTLADTSPEYAVRYTLDRHNRLOHW--DPATRKTYPNEPEAKLWK 509  
 DB 425 DKEALAQIQNI-----EMDYILKNADVEKAAAMRLRALVYKSW-----KIMR 465

## RESULT 32

Q8YPH1 PRELIMINARY; PRT; 480 AA.  
 AC Q8YPH1;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN ALR4223.  
 OS *Anabaena* sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriuch M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsui A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003595; BAB75922.1;  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; Pldc; 2.  
 DR SMART; SM00155; Pldc; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 480 AA; 54495 MW; 4F9C4130B5CBF789 CRC64;

Query Match 9.3%; Score 253; DB 16; Length 480;  
 Best Local Similarity 24.1%; Pred. No. 8.3e-11;  
 Matches 111; Conservative 68; Mismatches 174; Indels 108; Gaps 16;

QY 68 PHTNGSIDTYLLNDHEAFAAARALIESAHSIDQIYIWRNDISGRLLFNLVYLAENG 127  
 DB 115 PFTSG--NNAKLLINGQOYTAAMLSAASANSYLLQSYIVVDKAGNEPKALIAKAG 173  
 QY 128 VVRLLDDNNTRGIDDLALDSHPNIEVRLNPFVLRKWRALGYLTFPRLNRHMK 187  
 DB 174 IRVYLIYDEIGSNKISRLYVKSLOKDYOVSAFHTTRGKNF-----QLNFRNHRK 225  
 QY 188 SEFADNRATILGGRNIGDEYFVGED--TVPADLILATGSVGEVSHDF--DRYWASH 242  
 DB 226 ILVVDGRTAFIGLNTISDEY--LGKNRILSPWRDTHMLEGPTVOSLQCCFIQDWYMAT- 282  
 QY 243 SAHNATRIISGNIGKGLAGNDESRALLRYREYQSPYVOKIOTGRIDMOSVOT 302  
 DB 283 -----RQVIDVWQVOP-----NMESDYT 301  
 QY 303 RLISDDPAKGLDRDRKRPPIAGRL--ODALKOPEKSVYLVSPYPTKSGTDALAKVQ 359  
 DB 302 ALVFPFGPA-----DKLK--ACKLFFVSAIQAOGRILAIIPYFPDDSTLTALKLAL 353  
 QY 360 DGIQVTVLINSLOATVAAVHSGVYKRPYLKAGIKVLELOPNHNAVPAKDKGLTGSSV 419  
 DB 354 RQVDVAILPN--RPDHLVLYLCSFSYTEMKATNIKILRYK--HGF----- 396  
 QY 420 TSLHAKTFIVDGKRIFGSFNLDPRSARLNTG--MGVVISPIAEMERTLADTSPEYAV 478  
 DB 397 --MHQVILLIDKEMAGVGVNLNDRSFNLFPEMGVANSQFVKSVEEMLOADLAAALAV 454  
 QY 479 RVLDRHNRLOHWDPATRKTYPNEPEAKLIRIAKILSL 519  
 DB 455 -----DFSDYERKYLMEFLAVAISSLL 476

## RESULT 33

Q9LAEB PRELIMINARY; PRT; 397 AA.  
 AC Q9LAEB;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE YWJE protein.  
 GN YWJE.  
 OS *Bacillus cereus*.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;  
 OX NCBI\_TaxID=1396;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 14579 TYPE STRAIN;  
 RX MEDLINE=20053637; PubMed=10589720;  
 RA Oksdal O., Gornall M., Purnelle B., Rose M., Iereclus D., Kolsto A.B.;  
 RT "Sequence analysis of three *Bacillus cereus* loci under *Picr*-regulated  
 RT genes encoding degradative enzymes and enterotoxin.";  
 RL Microbiology 145:3129-3138(1999).  
 DR EMBL; AF243712; CAB69815.1;  
 DR InterPro; IPR001736; PLD.  
 DR Pfam; PF00614; Pldc; 2.  
 DR SMART; SM00155; Pldc; 2.  
 SQ SEQUENCE 397 AA; 46008 MW; 70F7D0044033EDA1 CRC64;

Query Match 9.3%; Score 251; DB 2; Length 397;  
 Best Local Similarity 22.3%; Pred. No. 8.7e-11;  
 Matches 108; Conservative 68; Mismatches 178; Indels 130; Gaps 14;

27 LCLLLSCSSMPL-PLF-----ERTSRHNTSKPVRLNLIIDIRTPHNGLSDIYL 78  
 4 LSLILISVALMIIITDLSGRLLHLKRVSSRTF---PLR-----OSDFHL 44  
 79 LNDPEAFARALIESAHSIDLOYIWRNDISGRLLFNLYLAERGVRRLIDNN 138  
 45 YTYGKDYDALFTDIKAQHNIHILFFIVKNDKISRFLKIIDKAQGIIEVRLIDFG 104  
 139 TRGL-DLLLLDLSHPT-----EYRLFPVYLKRMALGILTDPRLNRHNSFTAD 192  
 105 SHLSNFAIRSLQKHGVSFSCCHKYKFPPLPF-----SANQRHNRKITVID 150  
 193 NRATILGRNIGDEYFKGDEPT-VEADDDILATGVSVEVSHDFPRYASAHNATLI 251  
 151 GKTGYIGFNGEYLGHNELGLMRDYLRLTGSGVDLQKFLIDMF----- 199  
 252 RSGNIGKGLQALGYNDETSRHALLYRETVEQSP--LYQKIQTGRIDMOSVQRLISDD 308  
 200 -----DDTKQMLDASLTFPKQNEPTLIHQFIPTDGAVLYQTHFLIN-- 242  
 309 PAKGLDRRKRPIAGRLQDALKQPEKSVYLVSFVFPYTKSGTDALAKLVODGIDVLT 368  
 243 -----AAKKEICIGTFEIPGKIMNALLKARERGVOITIV 279  
 369 NSLQATDVAAVHSGVYKRPPLKAGIKLYELOPNHAPVAPKDKGLTSSVTSIAKTFI 428  
 280 P--EKADHPLVREAKFPYCRKLQAGCNIVAFQGF-----FIAKIII 320  
 429 VDGRRIFIGSNLDPRSARLNTENGVIIESPKIAEQMERTLADSPYAYRVLTDRNRL 488  
 321 VDDICIDIGTANFDMRSILYINHEINCLLYNKHFIQTQYK-----KFHEDIDNASVL 371  
 489 QWHD 492  
 372 SYSD 375

RESULT 34  
 08Y266 PRELIMINARY; PRT: 422 AA.  
 AC 08Y266; DT 01-MAR-2002 (Tremblrel, 20, Created)  
 DT 01-MAR-2002 (Tremblrel, 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)  
 DE Hypothetical protein RSC0470.  
 GN RSC0470 OR RS04425.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 NC Ralstonia.  
 OX NCBI\_Taxid=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GMT1000;  
 RX MEDLINE-21618179; Pubmed-11823852;  
 RA Salanoubet M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 Atilat M., Billault A., Brothier P., Camus J.C., Catolico L.,  
 Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
 Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,  
 Stiguer P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 EMBL: AL646059; CADI3998.1;  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLD; 2.  
 DR SMART: SM00155; PLD; 2.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 422 AA; 46969 MW; 5D0FBFC2AC9C40B6 CRC64;

Query Match 9.2%; Score 250.5; DB 16; Length 422;  
 Best Local Similarity 24.6%; Pred. No. 1e-10;  
 Matches 105; Conservative 65; Mismatches 191; Indels 65; Gaps 12;

76 IYLLNDPEAFARALIESAHSIDLOYIWRNDISGRLLFNLYLAERGVRRLID 135  
 25 IDLLHGGAFFPALIEAIDARRIALETYIYIDDDTGRVTEALARAAGVDRVRYT 84  
 136 DNNRGLDLDLLADSHNIEVRLFPN--FYLKRMALGYLTDPRLNRHNSFTAD 192  
 85 GFGTITLPAAGIAAMDAAGAMRYRPRGRFLOR-----RYLRUHRKVAVID 133  
 193 NRATILGRNIGDE--YKVGEDYFA--DDILATGVSVEVSHDFPRYASAHNAT 248  
 134 DEVAEVGGIINIIDDINHPRDDALGARYPDRAVRGRFLVQIALMVDRLM----- 184  
 249 RIIRSGNIGKGLQALGYNDETSRHALLYRETVEQSPLYQKIQTGRIDMOSVQRLISDD 308  
 185 --WQTGLLA-GVREYGVGVAEPFV-----TDP-----KPRRGASGTQHDQAPDN 230  
 309 PAKGL---DRDRKRPIAGRLQDALKQPEKSVYLVSFVFPYTKSGTDALAKLVODGIDVT 365  
 231 VLASLVLDGNVRRNRAIREYLRALGTARHEVYIANAFLPGVFRMRLAACRRGVYR 290  
 366 VLINSLOA--TVAAVHSGVYKRPPLKAGIKLYELOPNHAPVAPKDKGLTSSVTSIHA 424  
 291 ML---LQGHVEYAVQHYATRSILYHMLLRDGVIEIHEY-----TASFLLHA 330  
 425 KTFIVDCGRIFIGSNLDPRSARLNTENGVIIESPKIAEQMERTLADSPYAYRVLTDR 484  
 331 KVAAYDGMATVGSNIDPFSLLAREANVAVMDAGVQGLRAALEAIAHARHPITLADA 390  
 485 HNRLOW 490  
 391 HAARPW 396

RESULT 35  
 09KRH2 PRELIMINARY; PRT: 484 AA.  
 AC 09KRH2; DT 01-OCT-2000 (Tremblrel, 15, Created)  
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN VCI670.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_Taxid=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE-20406833; Pubmed-10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 Gao S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dai J., Sellers P.,  
 McDonald L., Uppertack T., Fleischmann R.D., Nierman W.C., White O.,  
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 cholerae";  
 RL Nature 406:477-483(2000).  
 EMBL: AE004245; AAF94820.1;  
 DR TIGR: VCI670;  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLD; 2.  
 DR SMART: SM00155; PLD; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 484 AA; 55358 MW; A576B56692C04393 CRC64;

Query Match 9.2%; Score 248.5; DB 16; Length 484;

Best Local Similarity 24.6%; Pred. No. 1.9e-10;  
Matches 110; Conservative 64; Mismatches 160; Indels 113; Gaps 20;

QY 57 RLDNIIQIR-HPHNTGSLDIYLLNDPHEAFARALLISAESHSLDQYIWMNDISGR 115  
DB 102 RIDELCNMNMGPALSG-NTLSLLNSPNEILHAIIDIERAFOJIMWEYIHPGLADA 160  
QY 116 LFNLYVLAERGVRVRLDDNNT-----RGDLDDLALDHPNIEVRLFN 161  
DB 161 VASAVIQAKRGYNVNLDDSDAGSPREFSRPEKMKRDKGI-EVVALEVP----- 211  
QY 162 PVLKRWALGYLTDPRLNRHMNKSFTADNRATLLGSRNIGD-EYFK---VGEDTYF 216  
DB 212 -----WRIE-----LRRLDLRQHKRIIYIDELIAYTGMNWDPAVFQONACVGQ---W 257  
QY 217 ADDLIATGSSVGEVS-HDFRYNASHAHNATRIIRSGNIGKGLQALGYNDETSRHAL 274  
DB 258 IDIMVKTGPYVNLVAISHICWD--WEFET-----GSRMLPKNPE----- 294  
QY 275 LRYRETVESPLYOKIOTGRIDMQSVOTRLISDDPAKGLDRDRKPPRIGRLQDALKOPE 334  
DB 295 CRLEPQOPHPI-QVYPSG-----PGMPENLISQ-----VLFLAINQAN 332  
QY 335 KSVYLVSPYFVPTKSGTDALAKLVODGIDVTYVLTNSLQATDVAHVHGYKRYKPKLLKAG 394  
DB 333 RSVCIITPFYFVSADLLATLTKMTAQRGIKYVDIIP--KKNDSLMOVMSRAFYGELLEAG 390  
QY 395 IKLYELQPHNHPATYDKDKLTGSSVTSLSAKTFIVDGKRIPIGSEFMDPRSARLNTMGV 454  
DB 391 VOIHEF-----DGLI-----LHRTKSVYIDQCLVGTVMIDMRSLMLNPELTL 433  
QY 455 VIESPKIAEOM---ERTLADTSPEYAY 478  
DB 434 AVDDLEFQOMHMLQOQYIDQSHSVY 460

RESULT 36  
Q98H18 PRELIMINARY: PRT: 487 AA.  
AC 098H18: 01-OCT-2001 (TREMBlrel. 18, Created)  
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
DE Cardiolipin synthase.  
GN MLR2852.  
OS Rhizobium loti (Mesorhizobium loti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAFF303099;  
RX MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT \*Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti.;  
RT DNA Res. 7:331-338(2000).  
DR EMBL; AP003000; BAB49878.1; .  
DR InterPro; IPR001736; PLD.  
DR Pfam; PF00614; PLDC; 2.  
DR SMART; SM00155; PLDC; 2.  
KW Complete Proteome.  
SQ SEQUENCE 487 AA; 52990 MW; A403155E9D51D562 CRC64;

Query Match 8.9%; Score 242.5; DB 16; Length 487;  
Best Local Similarity 24.2%; Pred. No. 5.4e-10;  
Matches 107; Conservative 51; Mismatches 148; Indels 137; Gaps 12;

QY 76 IYLLNDPHEAFARALLISAESHSLDQYIWMNDISGRLLFNLYVLAERGVRVRLD 135

DB 124 IDVLEGTDEAYAMCAIDGERSVLEETIIFPDNAVGLLEVESLAGAVRGYTVRLD 183  
QY 136 DNNTGRLDILLALDSDHPNIEVRLNFPVLRKRALGYLTDPRLNRHMNKSFTADNRA 195  
DB 184 AVGARSVSPTLGLHRRANIPADVFNIGIIGLR-----LRYANLRTHRKILVVDGYV 236  
QY 196 TILGGRNI-----GDEYFKYGEDTVFADLIATGSSVGEVSHDFRYMAS 241  
DB 237 AFTGNNINIKGSAEFGSSNSARDIFKVTGPAY-ADL-----FSVAE---j-DWRFAT 285  
QY 242 HSA--HNATRIIRSGNIGKGLQALGYNDETSRHALRYRETVESPLYOKIOTGRIDMQS 299  
DB 286 NEALKGDAMRI-----ATL-----SPAPGP 306  
QY 300 VOTRLISDDPAKGLDRDRKPPRIGRLQDALKOPEKSVYLVSPYFVPTKSGTDALAKVQ 359  
DB 307 MLYRAVASGPDASNETNHR-----LLIGAFSVAARSIRLMSYPFLDRELLISALTITAR 360  
QY 360 DGIADVLTNSLQATDVAHVHSGYVRYKPKLLKAGIKLYELQPHNHPATYDKDKLTGSSV 419  
DB 361 RGYEIDVY-----VAVNNLFLVDRM 382  
QY 420 TSL-----HAKTFIVDGKRIPIGSEFMDPRSARLNTMGVIESPK 460  
DB 383 TAFQDGLKNYCRIMWTEGPEHSHKSLSDGVAVYSSNLSARSLRNLPEIDLEVLDA 442  
QY 461 IAPQERTLADTSPEYAYRTLD 483  
DB 443 FAREIEARIG-SAIETAIPTVLD 464

RESULT 37  
Q9K251 PRELIMINARY: PRT: 395 AA.  
AC 09K251: 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE Putative phospholipase.  
GN SC07081 OR SC344.07C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Brown S.P., Harris D.;  
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Kendenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT \*A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb streptomycetes coelicolor A3(2) chromosome.;  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Godle A., Hidaigo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Metzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.,  
 RT "Complete genome sequence of the model actinomyete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL: AL354616; CAB89753.1;  
 DR InterPro: IPR001736; PLoD.  
 DR Pfam: PF00614; PLoD; 2.  
 DR SMART: SM00155; PLoD; 1.  
 SQ SEQUENCE 395 AA; 44541 MW; 7D2021762EEB697 CRC64;

Query Match 8.9%; Score 241; DB 16; Length 395;  
 Best Local Similarity 23.3%; Pred. No. 5e-10; Indels 102; Gaps 18;

Matches 104; Conservative 65; Mismatches 175; Indels 102; Gaps 18;

QY 57 RLNDLRIQHPPTNGLDIYLNDPHEAFARAALIESAHSLDQYXI-WRNDISGR 115  
 DB 21 RLRLIGVAATGEG---NELVALRNGDELFPMALGAIKRAEHTIDMTFVYKR---GQI 72  
 QY 116 LFNLYLAER---GVRRLLDNNNTGGLD-DLLALDSHPNIEVRLNPFVLRKRAL 171  
 DB 73 ARDEAALADARRSGVRRLLDGFAGKEIEDLDAMEA-AGVOIAWF-----RKP 123  
 QY 172 GLTDPRLRRHNSFTADNRATILGGRNIGDEYF---KVGEDYFADILATGSV 227  
 DB 124 LMLSF-KONRCHKALVIDEHTAFGTGVAIEEMCGDARGPGE---WRDTHVVRGGA 179  
 QY 228 VGEVSHDEDRYASHAHNATRIIRSGNIGKQALGYNDSTSHALLRYETVQSP 287  
 DB 180 VDGVAAAFQNMWA--ECHD-----ELYDRDRSDH-----TQGTSTIV 216  
 QY 288 QKIO-TGRIWQSVOTRLISDDPAKGLDRRRKPIAGRLDALKEPSVYLVSPYVP 346  
 DB 217 QVGRSGASFQMDQMT-----LIRVMTLSAERFLATAYFAP 254  
 QY 347 TKSGLDALAKLVODGIDVTLSLQATDVAHVSGYVKRPLKAGIKYLEOPNAV 406  
 DB 225 DTFYDLICATARRGVYIILPGHTQACQLAGQYHYR-LIDACVSTIEYOP----- 309  
 QY 407 PATKDKGLTGSSVSLAKFTIVDGKRIFGSFNLDPRSARLNTMGVAVIESPKIAE--- 463  
 DB 310 -----TWMHAKITVDGLALICSTNFRSRMDEIMLAVLDQETNGLD 356  
 QY 464 -----QWERTLADTSPEIAYRVTLDR 484  
 DB 357 RQFDADERSFAIEPTFRWKRRATLRR 382

## RESULT 38

08ZE12

ID 08ZE12 PRELIMINARY; PRT; 486 AA.

QY 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cardiolipin synthetase (EC 2.7.8.-).  
 GN CLS OR YFO2188.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;  
 RX MEDLINE=21470413; Pubmed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titchell R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarrega A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Leatwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Ratlifford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague."  
 RL Nature 413:523-527(2001).  
 DR EMBL: AJ414151; CAC90995.1;  
 DR InterPro: IPR001736; PLoD.  
 DR Pfam: PF00614; PLoD; 2.  
 DR SMART: SM00155; PLoD; 2.  
 KW Transferrase; Complete proteome.  
 SQ SEQUENCE 486 AA; 55127 MW; 5B0C78CD0A2D499 CRC64;

Query Match 8.8%; Score 239; DB 16; Length 486;  
 Best Local Similarity 22.0%; Pred. No. 9.9e-10;

Matches 121; Conservative 77; Mismatches 189; Indels 162; Gaps 23;

QY 17 LMKTRSLSLCLLC-----SCSSMLPPLPEE 43  
 DB 28 LMKRRAPVSAMAWLLIYLPVGLIAYLSFGLHGRRAERAKAMPSTARWLSLK- 86  
 QY 44 KRESRHF--NRSKPYRLNDIIRTPHTNGL--SDIYLLNDPHEAFARAALISAEHS 99  
 DB 87 --ECQHIFANSSEVASPLFCERRQINGVKGQLOLTTTDDTLALAVRDIABHN 144  
 QY 100 LDLYQYWRNDISGRLENL--VYLAERGVRLDLDNNT-GL 142  
 DB 145 IEMVFIYQ---PGSLVDQVVAESLMAAARGVCHCLLDLSAGSKQFFRSPPAMRNAGI 201  
 QY 143 DDLIALDSHPNIEVRLNPFVLRKRALGYLTDPRLRRHNSFTADNRATILGGRN 202  
 DB 202 EVV-----EALKVNVFRMFL-----RRMDLQHRKIVLIDNVAVATGSMN 241  
 QY 203 IGDEYFKYGEDYFADILIA-----TGSVGEVSHDEDRYASHAHNATRIIT---RSG 254  
 DB 242 MYDPREFKQDAGVGQIDMMAMMEGPAVATLGIY-----YACDWEIETGRIILPPDDA 295  
 QY 255 NIGKGLQALGYNDSTSHALLRYETVQSPYOKIQTRIDWQSVOTRLISDDPAKGLD 314  
 DB 296 NI-----MPFEETG-HTI-----QVIASGP--GFP 318  
 QY 315 RDRKRPPIAGRLDALKEPSVYLVSPYPTSGTDALAKLVODGIDVTLSLQAT 374  
 DB 319 EEM-----IHOALTFVAVAREQLIMTTFYFVPSDDLHAICTAQRGVASIIYV--REN 372  
 QY 375 DVAAVHSGVYVKRRLKAGIKIYELQPNHNAVPARKDKGLTGSSVTSIHAATFYDGRI 434  
 DB 373 DSMVWRWASRAFFTELLNAGVAKIYF-----EGGL-----LHRSKVLVDGOLS 415  
 QY 435 FIGSEFLDPRSARLNTMGVAVIESPKIAEOMERTLADSPPEAYAVTL---DRHN-RLQW 490  
 DB 416 LVGTYNLDMRSLIMLNFELTIVLD---DGFADLAQYQDDYIARSLDDEBRNKRPLW 471  
 QY 491 HDPATKTY 499  
 DB 472 HRYTERLFY 480

## RESULT 39

09CGC3

ID 09CGC3 PRELIMINARY; PRT; 487 AA.

QY 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN CLS OR LL1174.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Lactococcus.  
 OX NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IL1403;  
 RX MEDLINE=21235186; Pubmed=11337471;  
 RA Boletini A., Winkler P., Manger S., Jallion O., Malarme K.,  
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;



RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*  
 RT *lactis* ssp. *lactis* IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 DR EMBL: AE006349; AK05272.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 487 AA; 56322 MW; A454621B0A308268 CRC64;

Query Match 8.8%; Score 238.5; DB 16; Length 487;  
 Best Local Similarity 22.3%; Pred. No. 1.1e-09;  
 Matches 99; Conservative 62; Mismatches 169; Indels 113; Gaps 14;

QY 43 ERESHNPNSKPVRLDNLQIRHPHTNGLSIDYILINDHFAAARAALIESAHSIDL 102  
 DB 104 ERLKRYFN-----NMERTPLCRGNKNVNYLNG-EDKFSALFDIDIKNAKDNIHV 151  
 QY 103 QYVIMRNDISGRLLFNLYLAERGVAVRLLDNNTRGIDLLALLDSHPNIEVRLFP 162  
 DB 152 EYVAFNDRKIGTAFRDLNLIKAKEGEVRYVDPWGCKTRKDFEKLPLE----- 199  
 QY 163 FVLRKRALGYLDF-----PLNRMMHKSFTADNRATILGCRNIGDEX----- 207  
 DB 200 -----EAGGKYTAFTSRNTLTKRTLNLHARRIYVIDQISWTGFGFNVDQYIYNSKK 253  
 QY 208 FKVGEDTVEADLDILATGSVYGEVSHDFDRYMAHSHAHNATRLIRSGNIGKGLQALGYND 267  
 DB 254 FGWWRPT-----HGRIVGTAAFGLOE-----TIRDMNV----- 282  
 QY 268 ETSRHALRYRETVESQSPLYOKI--QTGRIDWQSVQTRLISDDPANGLDRRKRPPIAGR 325  
 DB 283 -----SVROPKDKLEREDSFYVVEEGNIDQ-----IYANGPKS--DKKTLRTGFIKM 330  
 QY 326 LQDALQPEKSVYLVSPYFPTKSGTDALAKLVQDGDIVYVLSLQATVAVAHSGYVK 365  
 DB 331 IMDA---EDYIWLQSPYLLPDDSMITLVAANSGVDVIMIPNM--PDHPIFRATQY 384  
 QY 386 YRRPLKAGIKLYELOPNHNAVPAATKDKGLGSSVTSLSHAKTFIVDGKRIFGSEFNDPRS 445  
 DB 385 YANYLKHGKIYNTNGF-----IHSKTLVMDGKLGVPGTINODIRS 427  
 QY 446 ARLNTEGVVIESPKIAEQMERT 468  
 DB 428 YELNFEISAFCYDETVAKESMRT 450

RESULT 40

Q8XNG8 PRELIMINARY; PRT; 511 AA.

AC Q8XNG8;  
 DT 01-MAR-2002 (TREMblrel. 20, Created)  
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Cardiolipin synthase.  
 GN CLS8 OR CPE0365.  
 OS Clostridium perfringens.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
 OC Clostritales; Clostridiaceae; Clostridium.  
 OX NCBI\_TaxId=1502;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=13 / TYPE A;  
 RX PubMed=11792842;  
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,  
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,  
 RT "Complete genome sequence of *Clostridium perfringens*, an anaerobic  
 RT flesh-eater."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).  
 DR EMBL: AP003186; BAB80071.1; -  
 DR InterPro: IPR001736; PLD.  
 DR Pfam: PF00614; PLDC; 2.  
 DR SMART: SM00155; PLDC; 2.

KW Complete proteome.  
 SQ SEQUENCE 511 AA; 59702 MW; 76B3BAEB4C2D019E CRC64;

Query Match 8.8%; Score 238; DB 16; Length 511;  
 Best Local Similarity 23.1%; Pred. No. 1.3e-09;  
 Matches 96; Conservative 61; Mismatches 166; Indels 92; Gaps 14;

QY 82 PHEAFAR-AALIESAHSIDLQYIWRNDISRLFNLYLAERGVAVRLLDNNTR 140  
 DB 153 PGEVFERLKEELKKAKHYIFMEFYHBEVMMNSITELIEEKVEGEVRYFYDDMGL 212  
 QY 141 G---LDDLLALLDSHPNIEVRLFPVLRKRALGYLDFP---RLNRMMHKSFTA 191  
 DB 213 GLTPYKYNEVLEA---KGICKMVFNEFV-----PLSLRMNNRDRKRTVI 255  
 QY 192 DNRATILGSRNIGDEYKVGEDTVEA---DLDLATGSVYGEVSHDFDRYMAHSHAHNAT 248  
 DB 256 DGHGTGFTGGINLADY--INEIVRFGHMKDASIMIKGDVAVNLTVMFLQIMNFEYSEGE- 312  
 QY 249 RIRSGNIGKGLQALGYNDETSRHALRYRETVESQSPLYOKIQOTGRIDWQSVQTRLISDD 308  
 DB 313 -----QEKYYPYFNHDEFEESDG---YQPYGDSPLDEIV-----GENYIANTIN-- 356  
 QY 309 PAKGLDRDRKRPPIAGRLQDALQPEKSVYLVSPYFPTKSGTDALAKLVQDGDIVTVLT 368  
 DB 357 -----KADYVYINPYYLIDNELVYALTLAKSGIDVRIYT 393  
 QY 369 NSLQATVAVAHSGYVYKRPPLKAGIKLYELOPNHNAVPAATKDKGLGSSVTSLSHAKTFI 428  
 DB 394 PHLE--DKWYAHIVTRAIVYQOLIESGVKIEYEPGF-----IHSKTFV 434  
 QY 429 VDGRKIFGSEFNDPRSARLNTENGAVVIESPKIAEQMERTLADTPEYAVRYVTD 483  
 DB 435 SDEIGIGVITNMDYRSLYLHFEQGVWLKXKTSVGOIKDDELKTL-EKQQRITILE 488

Search completed: May 12, 2003, 09:59:32  
 Job time : 99 secs



GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 09:55:41 ; Search time 46 Seconds  
(without alignments)  
1097.186 Million cell updates/sec

Title: US-10-066-551-4

Sequence: 1 MRANPQTQAMPSETISLMKT.....KIMKRIAKILSLPIESLL 525

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database: PIR\_73:\*

1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2713	100.0	525	2	B81859
2	2599	95.8	508	2	B81083
3	1115.5	41.1	493	2	C64847
4	1113.5	41.0	493	2	H90806
5	1106.5	40.8	493	2	D85666
6	1100	40.5	495	2	AE0636
7	821	30.3	518	2	G97556
8	821	30.3	518	2	AB2777
9	808.5	29.8	529	2	F82983
10	632	23.3	502	2	E71963
11	628	23.1	502	2	F64543
12	427	15.7	652	2	B82724
13	325	12.0	482	2	G70063
14	318	11.7	504	2	A11433
15	311	11.5	504	2	A11075
16	308.5	11.4	494	2	F90001
17	293.5	10.8	481	2	H66744
18	293.5	10.8	482	2	A11762
19	288.5	10.6	482	2	AG1387
20	281	10.4	476	2	T43863
21	275	10.1	505	2	G89906
22	274.5	10.1	490	2	B82971
23	274	10.1	503	2	C84125
24	273.5	10.1	500	2	S60089
25	264.5	9.7	492	2	AE3539
26	262.5	9.7	467	2	H82711
27	261	9.6	359	2	D83103
28	258	9.5	510	2	E97307

30	253	9.3	480	2	AH2333	cardiolipin synthase
31	248.5	9.2	484	2	C82171	cardiolipin synthase
32	239	8.8	486	2	AG0266	cardiolipin synthase
33	238.5	8.8	487	2	F86771	cardiolipin synthase
34	231	8.5	398	2	S53419	cardiolipin synthase
35	231	8.5	401	2	B83377	cardiolipin synthase
36	229	8.4	413	2	A10598	probable phospholipase
37	227	8.4	492	2	B97658	cardiolipin synthase
38	227	8.4	492	2	AC2882	cardiolipin synthase
39	222.5	8.2	532	2	G82872	cardiolipin synthase
40	222	8.2	413	2	C90737	cardiolipin synthase
41	222	8.2	413	2	D85587	cardiolipin synthase
42	222	8.2	413	2	E64815	cardiolipin synthase
43	219	8.1	474	2	B95023	probable phospholipase
44	219	8.1	510	2	D97894	cardiolipin synthase
45	216	8.0	486	2	AF0651	cardiolipin synthase

## ALIGNMENTS

RESULT 1  
B81859  
phospholipase D-family protein NMA1646 (imported) - Neisseria meningitidis (strain Z24  
C/Species: Neisseria meningitidis  
C/Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C/Accession: B81859  
R:Parhill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491  
A:Reference number: A81775; MUID:20222556; PMID:10761919  
A:Accession: B81859  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-525 <PAR>  
A:Cross-References: GB:AL162756; GB:AL157959; NID:97380091; PIDN:CA84874.1; PID:9736  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA1646

Query Match	Best Local Similarity	Score	DB 2:	Length	525:
Matches: 525; Conservative	100.0%;	Score 2713;	DB 2:	Length 525;	
		Pred. No. 1.3e-191;			
		Mismatches 0;	Indels 0;	Gaps 0;	
QY 1	MRANPQTQAMPSETISLMKTRSLISLCLLSCSSWLPPLERERESRHFNTSKPYRLDN 60				
DB 1	MRANPQTQAMPSETISLMKTRSLISLCLLSCSSWLPPLERERESRHFNTSKPYRLDN 60				
QY 61	ILQIRTPHTNGLSIDYLLNDPHEAFARAALIESAHEGLDQYIWRNDISGRLLFNLY 120				
DB 61	ILQIRTPHTNGLSIDYLLNDPHEAFARAALIESAHEGLDQYIWRNDISGRLLFNLY 120				
QY 121	YLAERGVRRLLDDNNTRGDLDDLALDSHPNIEVRLENFVLRKRALGYLDEPRL 180				
DB 121	YLAERGVRRLLDDNNTRGDLDDLALDSHPNIEVRLENFVLRKRALGYLDEPRL 180				
QY 181	NRMHNSFTADNRATILGGRNIGDEYFVAGDYLATGSGVGEVSHDEDRYVA 240				
DB 181	NRMHNSFTADNRATILGGRNIGDEYFVAGDYLATGSGVGEVSHDEDRYVA 240				
QY 241	SHSANAATRIIRSGNIGKGLAGYNDERSHALLRYRETVOSPLXOKIQGRIDMOSV 300				
DB 241	SHSANAATRIIRSGNIGKGLAGYNDERSHALLRYRETVOSPLXOKIQGRIDMOSV 300				
QY 301	QTRLLSDDEPAKGLDRDRKPPYAGRIODALKOPERSVYLVSPYPTKSGTDAKLKLYOD 360				
DB 301	QTRLLSDDEPAKGLDRDRKPPYAGRIODALKOPERSVYLVSPYPTKSGTDAKLKLYOD 360				
QY 361	GIDVTVLNSLOATQVAAVHSGYVYKRPPLAKAGIKLYELQPNHNAVPAATKDKGLTSSVT 420				
DB 361	GIDVTVLNSLOATQVAAVHSGYVYKRPPLAKAGIKLYELQPNHNAVPAATKDKGLTSSVT 420				



QY 421 SLHAKFTIYDVKRIIFIGSFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRV 480  
DB 421 SLHAKFTIYDVKRIIFIGSFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRV 480  
QY 481 TLDRHNRLOMHPDTRKTYPNPEAKLMKRIAKIISLPIESL 525  
DB 481 TLDRHNRLOMHPDTRKTYPNPEAKLMKRIAKIISLPIESL 525

RESULT 2

B81083  
cardiolipin synthetase family protein NMB1434 [imported] - Neisseria meningitidis (strain C:species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: B81083  
R:Retelink, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A. Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Donnelly, B.A.; ri, H.; Olin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizze, M. Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: B81083  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <EXT>  
A:Cross-references: GB:AE002494; GB:AE002098; NID:97226684; PIDN:AAFA1795.1; PID:9722667 A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1434

Query Match 95.8%; Score 2599; DB 2; Length 508;  
Best Local Similarity 98.8%; Pred. No. 3.1e-183;  
Matches 502; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 18 MKTRSLISLCLLSCSSWLPLEERTESRHNTPKPYRLNLIQIRHPHNGISDLY 77  
DB 1 MKTRSLISLCLLSCSSWLPLEERTESRHNTPKPYRLNLIQIRHPHNGISDLY 60  
QY 78 LNDPHEAFARALIESAHSIDLYIWRNDISGRLEFNLVLAERGVRLLDN 137  
DB 61 LNDPHEAFARALIESAHSIDLYIWRNDISGRLEFNLVLAERGVRLLDN 120  
QY 138 NTGSLDLDLALDHSNIEVRLNPNFVLRKRALGYLTDPRLNRKHNKSFADNRATI 197  
DB 121 NTGSLDLDLALDHSNIEVRLNPNFVLRKRALGYLTDPRLNRKHNKSFADNRATI 180  
QY 198 LGRNIGDEFFKYGEDTVPADLDILATGSVGEVSHDFDYMASSHANATRIIRSGD 257  
DB 181 LGRNIGDEFFKYGEDTVPADLDILATGSVGEVSHDFDYMASSHANATRIIRSGD 240  
QY 258 KGLQALGYNDTSRHALRLRYRETVESPLVOKIQTGRIDMOSVOTRLISDPAKGLDRDR 317  
DB 241 KGLQALGYNDTSRHALRLRYRETVESPLVOKIQTGRIDMOSVOTRLISDPAKGLDRDR 300  
QY 318 RKPPINGRLODALKOPEKSYLVSPYFVPKSGTDALAKLVODGIDVTYVTLNSLOATDVA 377  
DB 301 RKPPINGRLODALKOPEKSYLVSPYFVPKSGTDALAKLVODGIDVTYVTLNSLOATDVA 360  
QY 378 AVHSGVYKRPPLKAGIKYELQPNHAPATKDKLGTSVSLHAKFTIYDVKRIIFIG 437  
DB 361 AVHSGVYKRPPLKAGIKYELQPNHAPATKDKLGTSVSLHAKFTIYDVKRIIFIG 420  
QY 438 SFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRVTLDRHNRLOMHPDTRK 497  
DB 421 SFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRVTLDRHNRLOMHPDTRK 480  
498 TYNPEPAKLMKRIAKIISLPIESL 525  
81 TYNPEPAKLMKRIAKIISLPIESL 508

C64847  
ymdc protein - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: C64847  
R:Baltner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M. A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12. A:Reference number: A64720; MUID:97426617; PMID:9276503  
A:Accession: C64847  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-493 <BLAT>  
A:Cross-references: GB:AE000206; GB:U00096; NID:91787282; PIDN:AACT4130.1; PID:91787. A:Experimental source: strain K-12, substrain MG1655  
C:Genetics:  
A:Gene: ymdc

Query Match 41.1%; Score 1115.5; DB 2; Length 493;  
Best Local Similarity 48.3%; Pred. No. 3.7e-74;  
Matches 224; Conservative 81; Mismatches 142; Indels 17; Gaps 5;

QY 72 GLSDIYLLNDPHEAFARALIESAHSIDLYIWRNDISGRLEFNLVLAERGVRR 131  
DB 37 GCGGLEPERSLDAFAARALIESAHSIDLYIWRNDISGRLEFNLVLAERGVRR 96  
QY 132 LLDNDNTEGDLDDLALDHSNIEVRLNPNFVLRKRALGYLTDPRLNRKHNKSF 191  
DB 97 LLDNDNTEGDLDDLALDHSNIEVRLNPNFVLRKRALGYLTDPRLNRKHNKSF 156  
QY 192 DNRATILGGRNIGDEFFKYGEDTVPADLDILATGSVGEVSHDFDYMASSHANATRI 251  
DB 157 DGAVTLGGRNIGDEFFKYGEDTVPADLDILATGSVGEVSHDFDYMASSHANATRI 216  
QY 252 RSGNIGK-----LQALGYNDTSRHALRLRYRETVESPLVOKIQTGRIDMOSVOTRL 304  
DB 217 --DVEGEMADRIELPASHNDAMTH---RLRMESSEPFNLVDDTLPIMAKTIL 269  
QY 305 ISDPAKGLDRDRRKPPINGRLODALKOPEKSYLVSPYFVPKSGTDALAKLVODGIDV 364  
DB 270 ISDPAKGLDRDRRKPPINGRLODALKOPEKSYLVSPYFVPKSGTDALAKLVODGIDV 329  
QY 365 TYVTLNSLOATDVAHSGVYKRPPLKAGIKYELQPNHAPAT--KDKGLGSSVTSLSH 423  
DB 330 TYVTLNSLOATDVAHSGVYKRPPLKAGIKYELQPNHAPAT--KDKGLGSSVTSLSH 389  
QY 424 AKFTIYDVKRIIFIGSFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRV 483  
DB 390 AKFTIYDVKRIIFIGSFNLDPRSAKLNTMGVIESPKIAEQMERTLADTSPEYAYRV 449  
QY 484 RHNRLOMHPDTRK--TYNPEPAKLMKRIAKIISLPIESL 525  
DB 450 RHNRLOMHPDTRK--TYNPEPAKLMKRIAKIISLPIESL 493

RESULT 4

H90806  
probable synthase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509;  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 01-Mar-2002  
C:Accession: H90806  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. gaasveta, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: H90806  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-493 <HAV>  
A:Cross-references: GB:BA000077; PIDN:BA034847.1; PID:913360888; GSPDB:GN00154 A:Experimental source: strain O157:H7, substrain RIMD 0509952



DB 120 NTRGLDLDLLADSHNIEVLEFNPFLKRWALGYLDFPRLNRRMHSFTADNRATI 180  
 QY 198 LGGNIDGDEYFVGEEDTVFADLLATGSGVGEVSHDFRYSASHANATRIISGNIG 257  
 DB 181 LGGNIDGDEYFVGEEDTVFADLLATGSGVGEVSHDFRYSASHANATRIISGNIG 240  
 QY 258 KGLALGYNDETSRHALRYRETVESQPLVYKIQIGRIDMOSVQRLISDDPAKGLDR 317  
 DB 241 KGLALGYNDETSRHALRYRETVESQPLVYKIQIGRIDMOSVQRLISDDPAKGLDR 300  
 QY 318 RRPPIAGRLODALKOPEKSVYLVSPFYFVTKGTDALAKLVODGIDVYLTNSLQATDVA 377  
 DB 301 RRPPIAGRLODALKOPEKSVYLVSPFYFVTKGTDALAKLVODGIDVYLTNSLQATDVA 360  
 QY 378 AVHSGYVYKRPPLKAGIKLYELOPNHAYPATKDKGLTSSVTSIAKTFIYDGKRIFTG 437  
 DB 361 AVHSGYVYKRPPLKAGIKLYELOPNHAYPATKDKGLTSSVTSIAKTFIYDGKRIFTG 420  
 QY 438 SFNLDPRSARLNTMGVYVIESPKIAQOMERTLADTSPEYAYRVTLDRHNRLOMHPATRK 497  
 DB 421 SFNLDPRSARLNTMGVYVIESPKIAQOMERTLADTSPEYAYRVTLDRHNRLOMHPATRK 480  
 QY 498 TYPNEPEAKLMKRIAKIISLPIESL 525  
 DB 481 TYPNEPEAKLMKRIAKIISLPIESL 508

RESULT 3  
 AA75751  
 ID AA75751 standard; Protein: 507 AA.  
 XX AA75751;  
 DT 21-MAR-2000 (first entry)  
 DE Neisseria gonorrhoeae ORF 987 protein sequence SEQ ID NO:2974.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KM antibacterial; gene therapy.  
 OS Neisseria gonorrhoeae.  
 XX  
 PN WO957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PE 30-APR-1999; 99WC-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC.  
 XX  
 DR WPI: 2000-062150/05.  
 DR N-PSDB; AA254513.  
 XX  
 PT Novel Neisseria polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics  
 XX  
 PS Claim 2; Page 1394; 1453pp; English.

CC AA253015 to AA254536, AA254537 to AA254615, and AA74253 to AA75941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 507 AA:  
 Query Match 95.0%; Score 2577.5; DB 21; Length 507;  
 Best Local Similarity 98.2%; Pred. No. 2.1e-226;  
 Matches 499; Conservative 4; Mismatches 4; Indels 1; Gaps 1;  
 QY 18 MTRSLISLCLLSCSSWLPLEERTESRHFNTSKPVRDNIQIRHPTNGLSDIY 77  
 DB 1 MTRSLISLCLLSCSSWLPLEERTESRHFNTSKPVRDNIQIRHPTNGLSDIY 60  
 QY 78 LINDPHEAFARALIESAHSLSLQYIWRNDISGRLLFNLYVLAERGVRLLDDN 137  
 DB 61 LINDPHEAFARALIESAHSLSLQYIWRNDISGRLLFNLYVLAERGVRLLDDN 120  
 QY 138 NTRGLDLDLLADSHNIEVLEFNPFLKRWALGYLDFPRLNRRMHSFTADNRATI 197  
 DB 121 NTRGLDLDLLADSHNIEVLEFNPFLKRWALGYLDFPRLNRRMHSFTADNRATI 179  
 QY 198 LGGNIDGDEYFVGEEDTVFADLLATGSGVGEVSHDFRYSASHANATRIISGNIG 257  
 DB 180 LGGNIDGDEYFVGEEDTVFADLLATGSGVGEVSHDFRYSASHANATRIISGNIG 239  
 QY 258 KGLALGYNDETSRHALRYRETVESQPLVYKIQIGRIDMOSVQRLISDDPAKGLDR 317  
 DB 240 KGLALGYNDETSRHALRYRETVESQPLVYKIQIGRIDMOSVQRLISDDPAKGLDR 299  
 QY 318 RRPPIAGRLODALKOPEKSVYLVSPFYFVTKGTDALAKLVODGIDVYLTNSLQATDVA 377  
 DB 300 RRPPIAGRLODALKOPEKSVYLVSPFYFVTKGTDALAKLVODGIDVYLTNSLQATDVA 359  
 QY 378 AVHSGYVYKRPPLKAGIKLYELOPNHAYPATKDKGLTSSVTSIAKTFIYDGKRIFTG 437  
 DB 360 AVHSGYVYKRPPLKAGIKLYELOPNHAYPATKDKGLTSSVTSIAKTFIYDGKRIFTG 419  
 QY 438 SFNLDPRSARLNTMGVYVIESPKIAQOMERTLADTSPEYAYRVTLDRHNRLOMHPATRK 497  
 DB 420 SFNLDPRSARLNTMGVYVIESPKIAQOMERTLADTSPEYAYRVTLDRHNRLOMHPATRK 479  
 QY 498 TYPNEPEAKLMKRIAKIISLPIESL 525  
 DB 480 TYPNEPEAKLMKRIAKIISLPIESL 507

RESULT 4  
 AA755452  
 ID AA755452 standard; Protein: 502 AA.  
 XX  
 AC AA755452;  
 XX  
 DT 24-JUN-1998 (first entry)  
 DE H. pylori ORF 02ae11612\_33203250\_c1\_51 secreted protein.  
 XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 XX  
 OS Helicobacter pylori.

